DOCUMENT RESUME

ED 405 946 PS 021 192

AUTHOR	Ramp, Eugene A., Ed.; Pederson, Colleen Smalter, Ed.
TITLE	Follow Through: Program and Policy Issues. Proceedings from the Conference on the Feasibility of Follow-Up Studies to the Follow Through Program (Washington, D.C., February 21, 1991).
INSTITUTION	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Apr 92
CONTRACT	300-87-0114
NOTE	136p.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Educational Change; *Educational Policy; Elementary Education; *Followup Studies; *Longitudinal Studies; Program Descriptions; Program Effectiveness; *Program Evaluation; *Research Problems; Teacher Education
IDENTIFIERS	Educational Issues; *Project Follow Through; Project Head Start

ABSTRACT

Participants in this conference urged that arrangements be made for collecting and preserving information about participants, personnel, and organizations involved in the Follow Through Program, and for conducting long-term review studies of the program. Papers presented at the conference were: (1) "A Brief History" (Robert L. Egbert and Marijane E. England), which presents a history of project Head Start and of the four phases of the Follow Through Program between 1967 and 1991 (contains 104 references); (2) "Follow Through Graduates Today" (Garry L. McDaniels), which discusses the value of conducting a follow-up study of the program's graduates; (3) "Lessons Learned" (Russell Gersten and Wesley Becker), which discusses issues relevant to educational change (contains 27 references); (4) "Implications for Teaching" (Mary M. Kennedy), which examines the implications of educational models in the Follow Through Program for teacher learning and education and for children's lives, and the problem of drawing inferences from follow-up studies; (5) "Follow-Up of Follow Through: What Use for Policy?" (Richard F. Elmore), which focuses on problems associated with follow-up studies of planned intervention programs, and on the relationship of a follow-up study to policy; (6) "Potential Benefits of Longitudinal Studies in Education" (Herbert J. Walberg), which relates findings of studies on the influences of educational and psychological environments on children's learning and development (contains 59 references); (7) "Strategies for Subject Identification, Location, and Interviewing" (Sherri Oden), which proposes strategies to be used in follow-up studies of the Follow Through program; and (8) "Recommendations" (Robert Egbert, Edgar Epps, and Garry McDaniels), which proposes that the Department of Education conduct a long-term review of the Follow Through Program (contains 25 references). (BC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CE .ER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 405 946

Follow Through

Program and Policy Issues

Edited by:

Eugene A. Ramp
Colleen Smalter Pederson
The University of Kansas

PS 02 1192

Procedings from:

The Conference on the Feasibility
of Follow-up Studies to
the Follow Through Program

Sponsored by:

U. S. Department of Education
Office of Educational Research and Improvement
Office of Research

February 21, 1991
Washington, DC

BEST COPY AVAILABLE

U. S. Department of Education
Lamar Alexander, Jr.
Secretary

Office of Educational Research and Improvement
Diane Ravitch
Assistant Secretary

Office of Research
Milton Goldberg
Director

April 1992

Prepared for the Office of Research, Office of Educational Research and Improvement, under purchase order number 90-0346 SBA: 3-87-1-1800/ED: 300-87-0114. Individuals undertaking such projects are encouraged to express freely their professional judgment. This report, therefore, does not necessarily represent positions or policies of the U. S. Department of Education, and no official endorsement should be inferred.

Contents

vii *Foreword*

1 A Brief History
Robert L. Egbert
Marijane E. England

35 Follow Through Graduates Today
Garry L. McDaniels

43 Lessons Learned
Russell Gersten
Wesley Becker

57 Implications for Teaching
Mary M. Kennedy

73 Follow-Up of Follow Through: What Use for Policy?
Richard F. Elmore

87 Potential Benefits of Longitudinal Studies in Education
Herbert J. Walberg

107 Strategies for Subject Identification, Location, and Interviewing
Sherri Oden

125 Recommendations
Robert Egbert
Edgar Epps
Gary McDaniels

Foreword

Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world. This report is concerned with only one of the many causes and dimensions of the problem, but it is the one that undergirds American prosperity, security, and civility. We report to the American people that while we can take justifiable pride in what our schools and colleges have historically accomplished and contributed to the United States and the well-being of its people, the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people. What was unimaginable a generation ago has begun to occur—others are matching and surpassing our educational attainments.

—*A Nation At Risk*
U.S. Commission on Excellence in Education
1983

By the year 2000, all children in America will start school ready to learn.

By the year 2000, the high school graduation rate will increase to at least 90 percent.

By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.

By the year 2000, U.S. students will be first in the world in science and mathematics achievement.

By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

By the year 2000, every school in America will be free of drugs and violence and will offer a disciplined environment conducive to learning.

—National Educational Goals
U.S. Department of Education
July, 1990

The language of *A Nation At Risk* clearly prefigured the goals that were agreed to in the President's Education Summit. Nevertheless, in the seven years between 1983 when *A Nation At Risk* was published and 1990 when the President's Education Summit agreed on the six goals for American education, we made little progress in any of the areas represented by the goals.

- Head Start can accommodate, for even one year, fewer than 50 percent of those children most in need of preschool education.
- Scores on the NAEP and other tests have improved only slightly.
- Dropout rates in many urban schools are 50 percent or higher.
- Teenage pregnancy and drug usage rates remain high.
- Delinquency and crime rates among youth and young adults have changed little.

There is a growing sense of crisis over the apparent inability of our education and economic systems to compete internationally. The U. S. Department of Education is sounding the alarm and developing a vision that will help turn things around. The President and governors established the national goals for education and the Department launched *AMERICA 2000* to ensure that the goals are achieved.

As a part of developing a vision for the future, the Department is looking at exemplary programs from the past. The idea is to see what can be learned to help better inform the planners in creating the most effective strategy for future educational change. One program currently under study is Follow Through. This small program

has been around for nearly 25 years and is a successful example of what can be done to encourage educational change at the local level. Part of the "War on Poverty," it is one only a few programs that have survived. Along with the much larger and better known Head Start and Chapter 1 programs, Follow Through fills a void between the two. The one place in our education system where most at-risk children seem to fall through the cracks is during the transition from Head Start into the early elementary grades. The Department's Office of Educational Research and Improvement (OERI) is examining this program to see if there are promising avenues for future research and policy development.

Some areas of current, intense national interest where Follow Through might help to inform planning and policy include the following:

- *The President plans to serve all eligible Head Start children within the next few years:* With more Head Start graduates entering schools in the future, it will be even more important to ensure that gains are not lost. Follow Through exists primarily to help make the transition from Head Start into the early elementary grades as smooth and successful as possible for children at-risk of failure in school and later life. Congress authorized the program in 1967 to "follow through" on Head Start preschool children when they entered school.

- *Head Start graduates are exhibiting a serious loss of effects during their transition from preschool into the elementary grades:* According to R. Sargent Shriver, director of the Office of Economic Opportunity at the time, "the gains made in Head Start are being crushed by the broken promises of first grade." Head Start gains are sometimes lost within the first year or two of school. Comprehensive Follow Through models have been designed to help ensure that this does not happen. These models include both instructional and social service components. By working closely with Head Start and the schools, Follow Through maintains and even enhances the gains made by children in Head Start.

- *To counteract the intergenerational transmission of poverty, we must focus programs on both the child and the parent:* It is now clear that to effectively counteract the transmission of poverty from one generation to the next, programs must focus on *both* the child and the parent. Research conducted by the U. S. Department of Education revealed that of four federal programs studied, Follow Through had the most comprehensive and effective parental and community involvement components.

- *Chapter 1 (ESEA) is being reauthorized:* The multi-billion dollar Chapter 1 program is up for reauthorization next year. After more than a quarter-century, results from this giant program appear equivocal. Furthermore, Chapter 1 does not reach down into the early elementary grades and almost never reaches out to Head Start. The Follow Through strategy may be one example of how to more effectively reach the millions of at-risk children served by Chapter 1. By its design Follow Through is a program to prevent problems rather than a remedial program that only addresses children's needs after a problem occurs.

- *The first National Education Goal is that all children in America will start school ready to learn by the year 2000:* A key element of Follow Through is to begin building bridges from home and preschool into the elementary grades one to two years before children are ready to enter school. Program developers learned over the course of the last twenty years that schools cannot wait for the at-risk child to arrive before taking action to increase their chances for success. It is often too late by the time they reach school.

- *The strategy behind AMERICA 2000 involves working with schools to help them bring about fundamental change in the way they educate children–to help them "break-the-mold":* Follow Through programs operate in much the same way the *AMERICA 2000* strategy recommends for achieving fundamental school change. Though on a much smaller scale, Follow Through has resulted in a revolution that affects one child at a time, one teacher at a time, and one school at a time. In some instances, it has resulted in system-wide changes that last long after program funding disappears.

- *The NASDC competition to create New American Schools–a concept similar to Follow Through:* President Bush asked businesses to launch a private initiative to help *jump-start* schools across this nation. The result was the New American Schools Development Corporation. Its goal is to support the design and establishment of new high-performance learning environments that communities across the country can use to transform their schools for the next generation of American children. The concept is strikingly similar to that of Follow Through in that both programs stress parent and community involvement, comprehensive services, and adaptability to a wide range of sites.

- *The movement toward magnet schools and distinctive schools through site-based management:* The Follow Through program has provided an opportunity for the development of more than 20 distinctive models for enhancing learning, and these models are well documented. Because few other tested models exist for making elementary schools truly distinctive, districts interested in setting-up a system of elementary magnet schools might look to Follow Through for ideas and assistance in providing a wide range of options from which to choose. And Follow Through's mandate has always been to implement models with considerable local input and choice making the models an ideal fit into today's movement toward site-based management.

The Follow Through program began in September 1967 with 40 pilot projects in response to short-term evaluations that suggested that the gains in Head Start were lost during the early elementary years. During the time of its largest funding, Follow Through was the most comprehensive, thoroughly planned, family oriented, large scale effort made by schools in the United States. The program enrolled as many as 84,000 children (1972) in 173 different projects in urban and rural locations throughout the United States. At one time, more than 20 colleges and universities, regional educational laboratories, research and development centers, and other non-profit organizations provided program assistance and teacher training in the role of program sponsor to individual projects.

BEST COPY AVAILABLE

Dr. Robert Egbert, director of the program when school started in 1967, left Follow Through in 1971, but has maintained an interest in the program and, on several occasions, has urged that follow-up studies be conducted. With partial support of a grant from the EXXON Foundation, a group of individuals who had been or are currently associated with Follow Through came together in Denver in June 1990 to explore the possibility of long term follow-up studies. During two days of discussions, this group encouraged that the effort be continued and provided guidance on steps to be taken.

Subsequent to the Denver meeting, the authors met in late July with Dr. Milton Goldberg, Director of Research in the OERI, and with Dr. Ted Sanders, then Under Secretary in the Department of Education, and Mr. Nelson Ashline, his assistant. From these discussions, the Department of Education decided to convene a small conference at which a set of papers would be commissioned, presented, and discussed. The conference was held on February 21-22, 1991, in Washington, D. C. The papers in this volume are those solicited for presentation at the conference.

Dr. Milton Goldberg, conducted the conference. Conference participants, in addition to the paper authors, included: a) four expert panel members–Dr. Lois-Ellin Datta, former director of Head Start evaluation; Dr. Edgar Epps, University of Chicago professor of education and frequent expert consultant to Follow Through; Dr. Walter Haney, Boston College professor and writer of early histories and analyses of Follow Through; Dr. Gene Ramp, University of Kansas professor and current Follow Through program sponsor; b) two members of the participants in the June 1990 Denver meeting–Dr. Alice Paul, University of Arizona professor and current Follow Through program sponsor, and Mr. Richard Feldman, Bank Street College program sponsor; and c) several Department of Education and Office of Management and Budget staff members.

In two days of presentations and discussions, authors, expert panel members, and conference participants were unanimous in urging that arrangements be made for:

- Collecting and preserving as much primary source information about Follow Through participants, personnel, and organizations as is feasible.

- Conducting a set of long term review studies about Follow Through.

Although there was not unanimity on exactly what those studies should be, those present were in agreement that the original evaluation should not be revisited. That is, the focus in the review should not be on a nationwide comparison of graduates from the different sponsored programs. Drs. Edgar Epps, Garry McDaniels, and Robert Egbert reviewed the papers and discussion and prepared recommendations for the OERI. Those recommendations for a Follow Through Review can be found in the last article of this book.

Robert L. Egbert
Eugene A. Ramp
Marijane E. England

A Brief History

Robert L. Egbert
Marijane E. England
University of Nebraska

In the late 1960s two federal programs—Follow Through and Head Start Planned Variation—were launched with high hopes and ambitious, albeit somewhat conflicting, objectives. Both reflected faith in compensatory education—the belief that the lives and success of children from deprived homes could be significantly enhanced if the children were offered special school programs at an early age.

It is difficult a decade later (1975) to reconstruct the hectic hopeful atmosphere of Washington in the early days of the War on Poverty when new programs tumbled out of the White House and the Congress in rapid succession, and idealistic government officials worked frantically to get them started and confidently looked forward to quick and visible results. The launching of the Head Start program was typical of the spirit of the era. . . . Those eager to find effective ways of improving the life chances of the poor felt that special preschool programs for deprived children would give them a head start, compensate for the vocabulary skills that middle-class children learned at home, and enable them to function more effectively as they moved through the public schools.

—*Planned Variation in Education*
Alice Rivlin and Michael Timpane
1975

These two partial paragraphs capture something of the circumstances surrounding the initiation of the Head Start (1965) and Follow Through (1967) programs. Each of these programs was a high visibility undertaking within Lyndon Johnson's War Against Poverty. Head Start has continued to thrive; Follow Through has almost disappeared.

The nineteen-fifties and -sixties were decades of turmoil in the United States. The 1954 Supreme Court decision in Brown vs The Topeka Board of Education gave the civil rights movement a popular legitimacy that it had not enjoyed previously. Six years later John F. Kennedy was elected president; his request, "Ask not what your country can do for you; ask what you can do for your country" expressed the idealism that many felt. During the next ten years, this theme, was expressed repeatedly throughout the country. The idealism suggested by the theme led to the March 16, 1964 initiation of the War Against Poverty which included a number of new programs, including Head Start and, two years later, Follow Through—the focus of this paper.

Background: Head Start

President Kennedy was assassinated on November 22, 1963, but the nation's idealism continued unabated, and a year later the Economic Opportunity Act (PL 88-452), a key in the War Against Poverty, was passed by Congress and signed by President Johnson. Title II of the Economic Opportunity Act (EOA) authorized Urban and Rural Community Action Programs. Head Start was begun under Section 207 in Title II of the EOA that allotted funds for Research, Training and Demonstrations.

Two Conceptual Confusions

Head Start was plagued almost from the first by two conceptual confusions. These confusions also became a part of Follow Through's heritage. The first confusion was whether Head Start was a Community Action Program or a child development program. By authorization and funding source, Head Start was a Community Action Program; by intent of those who designed it, Head Start was a child development program. That is, it was planned and conducted largely by a combination of early childhood educators and health care professionals who, in 1965, were much more concerned with the child than they were with the community in which the child lived. Thus, they focused the program on the child.

Despite Head Start's popularity with the public many of those persons responsible for designing and administering community action programs, under which resources used to support Head Start were appropriated, were critical of it. They maintained that Head Start did not provide adequately for community control of its projects (e.g., Levin, 1967) and that funds drained off by Head Start should be used for real community action projects, i.e. projects that were controlled by members of the target community. In fact, some Head Start projects were legitimate community action programs. The Child Development Group of Mississippi (CDGM), for instance, was quite successful in its community action functions (Levin, 1967; Greenberg, 1969), so successful, in fact, that Head Start ultimately funded a competing program, Mississippi Action for Progress (MAP), to satisfy the concerns of the more establishment-oriented people of Mississippi. Most of the concerns of the community

action group ultimately were accommodated by the formation of Head Start Policy Advisory Committees. Eventually "advisory" was dropped from the title. These committees controlled what took place in Head Start projects, including hiring of personnel, recruiting children into the project, and forming local policies.

The concern of community action persons became important to the Follow Through Program too because it had the same funding source as Head Start. The Office of Economic Opportunity insisted that local community action programs have a sign-off on all Follow Through projects and that all Follow Through projects have policy advisory committees. Like Head Start policy committees in their projects, these committees were expected to make important decisions in Follow Through projects. Some districts refused to accept Follow Through funding because school boards did not want poor parents making decisions.

The second misunderstanding that created problems for Head Start and that carried over to Follow Through centered around whether Head Start was a child development program or an early academic program. The intent of those who designed the program was clear. Head Start was to be a broadly conceived service program that was concerned with the child's physical health and abilities, emotional and social development, self-confidence in future learning efforts, and capacity to relate to others—and that Head Start should provide parallel opportunities for the family (Davens, June 19, 1968). In fact, the Office of Economic Opportunity formed a panel, headed by Robert Cooke, that prepared the following set of goals for Head Start.

- Improving the child's physical health and physical abilities.
- Helping the emotional and social development of the child by encouraging self-confidence, spontaneity, curiosity and self-discipline.
- Establishing patterns and expectations of success for the child which will create a climate of confidence for his future learning efforts.
- Increasing the child's capacity to relate positively to family members and others while at the same time strengthening the family's ability to relate positively to the child and his problems.
- Developing in the child and his family a responsible attitude toward society, and fostering constructive opportunities for society to work together with the poor in solving their problems.
- Increasing the sense of dignity and self-worth within the child and his family (Kirschner Associates, Inc., 1970, p. 21).

Although Head Start was created with this set of goals, it often has been considered a program intended primarily to advance children on early measures of academic progress. The earlier quotation from Rivlin is an example of this orientation. "Those eager to find effective ways of improving the life chances of the poor felt that special preschool programs for deprived children would give them a head start, compensate for the vocabulary skills that middle-class children learned at home, and enable them to function more effectively as they moved through the public schools" (Rivlin and Timpane, 1975, p. 3). It was partly the lack of information about the broad purposes of Head Start, as well as assumptions concerning the predictors of school success, that resulted in the study which led to Follow Through being initiated.

The Follow Through Decision

The decision to request a Follow Through program resulted largely from a single report that was based on beginning- and end-of-kindergarten data for children in four schools in New York City (Wolff and Stein,1966; Wolff and Stein, 1967). The authors concluded that gains which children made on achievement measures in the first summer of Head Start disappeared during the ensuing school year.

When they first reported their study at the annual meeting of the American Psychological Association in 1966, Wolff and Stein created considerable consternation among those who supported Head Start. The enthusiasm which these persons felt for preschool programs for children, especially children who were being raised in poverty, had been fueled by Head Start's first summer enrollment. This program which had been planned as a relatively modest pilot effort enrolling not more than about 100,000 children in the summer of 1965 had brought about a ground swell of interest that spread across the country until 560,000 children in almost fifty percent of the nation's counties were enrolled (McDavid, Gordon, Grotberg, and Datta, 1968). Furthermore, Head Start advocates were encouraged in their enthusiasm by both consultant reports and the preliminary results from the first studies conducted on the program. Taken together, the enrollment, consultant reports, and first results suggested that Head Start should be expanded and extended. The Wolff and Stein findings, on the other hand, created doubts in some persons' minds about the program, especially any expansion of it.

Two months after Wolff and Stein reported their Head Start findings, Sargent Shriver, in his address to the Great Cities Research Council, gave the Office of Economic Opportunity response to that report. Shriver did not criticize the Wolff and Stein study for being narrowly conceived, nor did he argue that a lengthened Head Start would make permanent the early achievement test gains. Instead, he stated, "The readiness and receptivity they (the children) had gained in Head Start has been crushed by the broken promises of the first grade." He called for a follow-up of Head Start children into the early elementary grades. The program for providing such follow-up would be called Follow Through (Egbert, 1973). President Johnson (1967) in his State of the Union Message and again in his February 8 message on Children and Youth, built on Shriver's recommendation with a formal request for a Follow Through program. "Head Start occupies only part of a child's day and ends all too soon. He often returns home to conditions which breed despair. If these forces are not to engulf the child and wipe out the gains of Head Start, more is required." (p. 37) Beginning with this argument, President Johnson moved forward with a request for a Follow Through Program. Thus, Follow Through was begun to preserve and build on the gains that children made in Head Start, gains that some persons began associating with academic achievement (Egbert, 1973).

Head Start and Follow Through Authorization

The early years of school were not mentioned in the initial (1964) version of the Economic Opportunity Act, but then neither were preschool programs. Both Head Start and Follow Through were authorized for the first time in the 1967 amendments to the EOA which were passed by Congress and signed by the President in December 1967. Head Start and Follow Through became (PL-88-452, Section 222[a], Paragraphs 1 and 2,1964 as amended through PL 90-222).

(1) A program to be known as "Project Head Start" focused upon children who have not reached the age of compulsory school attendance which (A) will provide such comprehensive health, nutritional, education, social, and other services as the director finds will aid the children to attain their full potential, and (B) will provide for direct participation of the parents of such children in development, conduct, and overall program direction at the local level.

(2) A program to be known as "Follow Through" focused primarily upon children in kindergarten or elementary school who were previously enrolled in Head Start or similar programs and designed to provide comprehensive services and parent participation activities as described in paragraph (1), which the director finds will aid in the continuing development of children to their full potential. Funds for such programs shall be transferred directly from the director (of OEO) to the secretary of Health, Education and Welfare. Financial assistance for such projects shall be provided by the secretary on the basis of agreements reached with the director directly to local educational agencies except as otherwise provided by such agreements.

Follow Through: Phase 1—1967

Long before Follow Through was actually authorized, the staff of the U.S. Office of Education (USOE) and the Office of Economic Opportunity (OEO) began to plan for the major operational program scheduled to commence in the fall of 1968. Three important actions were taken. (1) Follow Through was to be administered in the Division of Compensatory Education (DCE) within the Bureau of Elementary and Secondary Education (BESE). (2) A Follow Through Advisory Committee with broad representation from early childhood education, the social sciences, and school administration was appointed to make recommendations on program content. (3) OEO transferred money to fund a pilot phase of Follow Through—$.3 million in FY 1967 and $2.5 million in FY 1968 funds (Egbert, 1973).

The President signed a Delegation of Authority for Follow Through on June 26, 1967. This delegation authorized the Department of Health, Education and Welfare—and subsequently the U. S. Office of Education—to administer the Follow Through pilot program. At the same time, a Memorandum of Understanding that governed the terms of the program administration was concluded by OEO Director Sargent Shriver and acting HEW Secretary, Wilbur Cohen (Egbert, 1973). The first Follow Through director was approved by both the Office of Economic Opportunity and the Office of Education (Shriver, July13, 1967; Howe, July 18, 1967).

Follow Through and Politics

Early in Follow Through's history, the political nature of visible government programs became apparent. By pre-arrangement, approximately 100 school districts were invited to apply for thirty pilot project grants. These projects were intended to develop an experiential base for the large scale Follow Through effort anticipated for the 1968-69 school year. Most of the invited districts applied. When the thirty winners were announced, certain key politicians objected to their constituents' projects

BEST COPY AVAILABLE

not being funded. Notable among these politicians were the congressman from Tampa, a senator from Rhode Island, and the mayor of Chicago. Not entirely unexpectedly, the Office of Education decided that funding an additional ten projects was possible, and staff members re-considered those projects not funded in the initial process. Also not entirely unexpectedly, Tampa, Providence, and Chicago all were among the ten new projects selected. Around the Division of Compensatory Education, this re-consideration became known as the Tampa Round of selection (Egbert, 1981).

From a financial perspective, this incident was relatively unimportant; experientially, it was critical. Once again, political expediency had been established as a vital feature in all important decisions.

Although administration of Follow Through was delegated to the United States Office of Education (USOE), OEO and Head Start personnel, representing OEO, retained a co-equal interest in the program. For example, on August 3, 1967, the Head Start director and his assistant met with the director of DCE, his assistant, and the director-designate of Follow Through to establish procedures for selecting Follow Through grantees for what was then anticipated to be a program serving up to 200,000 children in the 1968-69 school year (Herzman, August 3, 1967). The group specified more than two pages of details to be observed in the selection process. Shortly after this, Head Start appointed a staff person (a one-year Washington Fellow) to serve as liaison with Follow Through. This person monitored the purity of Follow Through and Follow Through local projects, particularly as that purity related to Head Start/Community Action Agency principles of parent control and comprehensive services to children (e.g. Orton, September 30, 1968; Egbert, October 14, 1968). On numerous later occasions (e.g., September 16 & 17, 1968 and November 18 & 27, 1968) senior Head Start personnel participated in planning meetings with senior Follow Through personnel staff and OEO made the final determination of what was required under the law (Egbert, 1981). For example, the twenty percent non-federal share contribution that was required under the EOA, Title II presented a serious problem to many school districts. OEO would not concede that Follow Through had any freedom of interpretation of this requirement (Boikess, May 21, 1968). (Non-federal share was a special problem to Follow Through schools for they could not count the use of building space as an in-kind contribution because it was where the children met anyway. On the other hand, because schools usually did not have a preschool program, school building space used rent-free by Head Start children could be counted as an in-kind contribution to the non-federal share in Head Start projects Egbert, 1981).

Reduced Budget

In mid-October 1967 rumblings began within the federal establishment to the effect that Follow Through would not be funded at $120 million (Hughes, October 18, 1967), but it was not until somewhat later that a semi-formal announcement was made that the most the program could expect for FY 1969 was $15 million. This amount, and others, were debated back and forth throughout the fall (Egbert, December 14, 1967). Not until after the first of the year (1968) did it become certain that Follow Through would receive any future funding. The amount turned out to be $15 million, less the $3.75 million borrowed for pilot projects in 1967-68 (Egbert, 1981).

Follow Through Program: Phase 2—1968-1972

The realization that Follow Through would not have funds for an operational program produced an atmosphere of haste and confusion. Discussions within DHEW, OEO and USOE led to the decision that Follow Through should be a program to produce information about how to work more effectively with children from low income families. Those who were responsible for designing the program response to this general decision had to plan a program that would permit producing such information; they also had to consider both the Follow Through legislation and the various interest groups, e.g., school districts, community groups, members of Congress, and professional groups, that would be concerned with the manner in which any large scale expansion of Follow Through was implemented.

In 1968, when Follow Through funding was reduced, the change from a service program to an experimental program was thought of as a temporary expediency. Notes from a Sept. 16-17, 1968 planning meeting that involved the USOE's Division of Compensatory Education Director, Head Start's Deputy Director, and the BOB (OMB) budget examiner responsible for Head Start and Follow Through contain several references to Follow Through becoming a service program (Follow Through Conference notes, 1968). The same concept was expressed in a June 19, 1969 memorandum from the Chief of OEO's Evaluation Division in RP/E to the Director of Follow Through (Evans, 1969) "...the time will soon arrive when some decision is going to have to be made about a major expansion of the Follow Through program. Indeed, such decisions are already being made."

From the relative calm of monitoring on-going projects, preparing guidelines, and planning a school-year 1968-69 program of reasonably well-defined parameters, Follow Through was plunged into the requirement, and the opportunity, of planning an experimental program of unknown nature and dimensions. Not only must the shape of the program be changed, the expectations of various, quite diverse constituent groups had to be acknowledged and dealt with.

Planned Variation

In order to provide a structure for projects and for consistency across them, Richard Snyder, director of Follow Through's Research and Evaluation Section, conceived a program model/program sponsor concept. The strategy of sponsorship required that each project (1) select from a set of pre-developed, pre-determined approaches the one they would like to adopt and (2) work with the program sponsor in the further development and implementation of the approach. The arrangement in which sets of local projects would work with program sponsors which they selected from those approved by Follow Through became known as planned variation (Egbert, 1981).

The planned variation concept was developed during the period from November 1967 through January 1968 in a series of formal and informal meetings, conversations, and telephone calls. Snyder and his assistant, Frieda Denenmark, talked with many of the nation's leading research child psychologists and early childhood program development specialists. Based on those discussions they arranged and conducted a series of four meetings.

The first meeting, December 18-19, 1967, was devoted to a discussion of the planned variation and model sponsor concepts. Though agreement was not unanimous, this meeting did serve to confirm the Follow Through staff judgment that the program sponsor concept was viable and should be implemented (Egbert, 1981).

The initial meeting was followed by two meetings of potential Program sponsors on January 5 and 6 and January 26 and 27 (Follow Through Meeting Notes, January 5/6 and January 26/27,1968; Egbert, Jan. 6, 1981). Whereas the December meeting had been largely limited to theoreticians/researchers, the January meetings were devoted to hearing from persons who either were established early childhood educators who represented a particular perspective, e.g. Elizabeth Gilkeson from Bank Street, William Hull from Educational Development Corporation and Marie Hughes from the University of Arizona, or were persons who were developing an apparently significant, theoretically based, new approach to educating young children. Included in this latter group were such persons as Ira Gordon (University of Florida) and Susan Gray (George Peabody College) who used a parent training approach and Siegfried Engelmann, (Illinois) Larry Gotkin (NYU) and Don Bushell (Kansas) who derived their programs from behavioristic psychology.

From the presentations at the January 1968 planning meetings, it was obvious that despite the growing interest in early childhood education and despite the extensive publicity given various new programs, no one was fully prepared to move into the primary grades with a completely developed, radically different approach to working with young children. For example, the highly publicized Engelmann approach (Bereiter & Engelmann, 1966) was partially developed for children's first educational experience—preschool or kindergarten—but not beyond.

Despite their limitations, a number of approaches seemed to be sufficiently well developed and to have a sufficiently secure and supportive institutional base that including them in Follow Through was justified. However, it became clear at this time that Follow Through sponsors would need to continue their program development efforts at the same time that they were working on implementation strategies and helping communities to begin their Follow Through projects.

The two January meetings of program sponsors were followed by one on February 9-11 during which an attempt was made to achieve working relationships between potential sponsors of similar persuasion and to secure some better understanding of what might be involved in program sponsorship. Neither effort was markedly successful. Because it was obvious that certain sets of approaches (models) were derived from common theoretical bases (e.g. Kansas, Illinois [later Oregon], and Pittsburgh all had behaviorism as their base) some time was spent in exploring whether such groups might form consortia or, at least, common interest groups. While common elements were recognized in these discussions, differences also were noted. For example, Kansas depended almost entirely on published materials and individual reinforcement procedures; Oregon and Pittsburgh produced their own, quite different materials, but Oregon utilized extensive small group, direct instruction, while Pittsburgh used individual materials and individual progress. Furthermore, there were strong institutional identifications. As a result, each sponsoring institution remained as a separate approach. Analyzing the nature and degree of program similarities and differences for the various sponsors became a major task in the national evaluation (Egbert, 1981).

BEST COPY AVAILABLE

Follow Through administrators did not view planned variation as a classic experiment. For example, in describing the program to an external review panel, the Follow Through director said, "By the time the appropriations bill had been passed, an essential agreement had been reached... that Follow Through would be a research and development program. Local programs would be funded as previously planned (and) an R & D program would be superimposed—an R & D (program) in which there would be deliberate variation of program approach, deliberate variation in contrast to the sort of variation that normally occurs in a local community when it develops its own program" (Egbert, May 14,1970, pp. 2-3). On the other hand, Alice Rivlin, who was HEW Deputy Assistant Secretary for Planning and Evaluation at the time that planned variation was designed, wrote seven years later, "It was never made clear to all concerned that these programs (Head Start Planned Variation and Follow Through) were planned variation experiments whose primary purpose was to try out and evaluate different approaches to early education . . .The shift in objectives was clear enough at the policymaking level (but it was not) made clear to many of the lower-level federal, state, and-local officials within whose gambit Follow Through was required to operate" (Rivlin and Timpane, 1975, pp.12-13).

The Sponsor Concept

The notion of program sponsorship has had different meanings for different people; furthermore, the meaning has shifted over time, both for sponsors and for others closely associated with Follow Through.

The simplest definition of program sponsor (approach) is an instructional model, fully developed, completely static, and uniformly implemented. A number of persons associated with program evaluation, appear to have identified with this definition. They appear to consider any program deviation from this definition to be a failure and any conceptual deviation to be wrong. They also tend to use the term "experiment", attaching it as a modifier either to Follow Through or to planned variation (Wisler, et al., 1978; Mosteller, 1975). In fact, by 1970, some of those persons with OEO responsibility for Follow Through became so disenchanted with Follow Through's lack of definition and control that they implemented a controlled experiment in which there were program models and randomly assigned schools. This experiment—performance contracting—met the technical requirements for an experiment which Follow Through did not, but the only outcome from that effort seemed to be that more than one year was required to "install" and test even math and reading models (Office of Economic Opportunity, February 1, 1972), a fact to which Follow Through sponsors gladly would have attested prior to the initiation of the Performance Contracting experiment.

The definition/description of sponsors which operated in Follow Through is that (a) the sponsor had developed a promising approach to working with young children, (b) the approach had a theoretical basis, (c) the sponsor was willing to work with a number of communities in implementing the approach, (d) the sponsor had a supportive institutional base, and (e) the sponsor accepted mutual accountability with the local project for the program's implementation and success.

By contrast, in the Performance Contracting experiment, the model did not need to describe a theoretical basis nor was there mutual accountability. Instead of mutual accountability, there were carefully prescribed separate responsibilities. The school

district was responsible to provide the students for a specified number of instructional periods and minutes; the contracting firm was responsible to provide instruction leading to achievement gains. The firm was to be paid by OEO in proportion to test score gains achieved beyond a certain minimum grade level increase (General Accounting Office, 1973).

Status of Follow Through Sponsors in 1968

Program sponsors that became associated with Follow Through in 1968 (or in succeeding years) were in varying stages of development. They ranged from the Bank Street College of Education approach that had been developed and tested over a period of several decades to a number of programs that were still in an early developmental phase and had scarcely been tried at all with school aged children, e.g. those approaches described by Engelmann and Becker, Bushell, Gotkin, Smock, and Hodges. However, each one had been written up in the professional literature and several had received publicity in the popular press, publicity suggesting an exciting, highly successful program (e.g. Newsweek January 29, 1968, pp. 47-48; Pines, 1966). Something of the tentative nature of the program status for some sponsors is evidenced by a quotation from the Engelmann-Becker proposal for participation in Follow Through. "The curriculum focuses on 3 major academic areas—language concepts, arithmetic, and reading. Programmed material will be available for the reading and perhaps for the language programs by the beginning of the fall semester" (Proposal for Bereiter-Engelmann Participation in Project Follow Through, undated mimeo p. i).

In his proposal, Bushell wrote about tokens given contingently, reinforcing events in the classroom, and programmed materials, but he did not describe how these would be put together in a Follow Through classroom (Bushell mimeo, undated). The reason that he did not describe a detailed program was that one had not been completely worked through. The group of behaviorists at The University of Kansas (Baer, Wolf, Risley, etc.) had published extensively concerning their work with young children and had established a continuing relationship with Head Start. Bushell, a new arrival in the group, had demonstrated that motivation and reading achievement could be improved through reinforcement techniques. It was because of this demonstrated competence in working with what was then a relatively new technology that had grown from behavioral psychology, rather than an established program, that caused Follow Through to invite The University of Kansas to submit a program sponsor proposal.

The Far West Laboratory for Educational Research and Development (FWLERD) proposed a program that was a "logical extension of the program that was developed for environmentally deprived three- and four-year-old children at the New Nursery School in Greeley, Colorado" (Nimnicht FWLERD mimeo, March 6, 1968). The New Nursery School was created by Glen Nimnicht, who also was to manage the FWLERD Follow Through sponsorship. It was based on the assumption that the child's self-concept, motivation, and learning skills as well as his/her achievement would be improved through working with equipment and materials that themselves gave feedback of success.

Clearly none of these three potential model sponsors had a fully developed program ready "to install and be evaluated." Instead, each one of them had a well described

theoretical base, a partial program and a clear notion of what a full program might be, good leadership, and a strong support institution. Several other groups were at a similar stage of development, e.g. those at the University of Florida and the University of Georgia (Egbert, 1973).

As indicated above, at the other end of the spectrum from those sponsor approaches that had come into being in the mid-1960s was the approach sponsored by the Bank Street College of Education. This approach to the education of elementary school children had its roots in Deweyian philosophy and analytic psychology and had been developed and refined during half a century of working in its own laboratory and in the New York City schools. Bank Street did not propose to do anything radically different in Follow Through from what it had already been doing (Egbert, 1973).

Somewhere between Bank Street and the new arrivals to early childhood education in its degree of program development was the University of Arizona. Dr. Marie Hughes, the initial manager of that sponsorship, had a life long history of work in elementary education. In the 1960s she initiated a project with the Ochoa Elementary School in Tucson which emphasized language competence and other skills the child would need in order to succeed in a changing and technical society. Special attention was given to the needs of non-English speaking children. As partial support for this project and related activities, Dr. Hughes secured funding as a Center in the National Laboratory for Early Childhood Education. From her extensive experience in working with elementary school age children and their teachers and her specific work at the Ochoa School, Dr. Hughes proposed to Follow Through a new, although well-developed, approach to early education (Egbert, 1973).

Although the proposed program was Follow Through's primary focus in considering a potential sponsor, also of great importance was the sponsor's ability to work effectively with a set of local communities in the adaptation and implementation of the program. The nature of program development was reasonably well understood; the issues involved in local adaptation and implementation were much less well understood, and there were few precedents in education for such an undertaking. Certainly, the traditional teacher education institution/local school district relationship did not provide very useful guidance. In the teacher education instance the teacher was educated in one institution and then employed by another on an individual basis. The teacher education institution had no responsibility for the school district's program, and the school district employed the teacher for his/her personal capabilities rather than because of any particular confidence in the institution granting the degree. There was no required institutional relationship involved. In Follow Through, on the other hand, the sponsor had a strong, continuing commitment to the nature and success of the district's program and, hence, to the performance of the individual teacher. The district was concerned about the sponsor's philosophical base because that base shaped the entire program. It also was concerned about the sponsor's ability to deliver its program (Egbert, 1973).

In order to be effective in transmitting (with appropriate local adaptation) a complex educational program, the Follow Through sponsor had to devise a delivery system which would both insure that the program's intent was properly implemented and that adequate feedback was provided to assist in making needed specific and generic modifications.

BEST COPY AVAILABLE

A number of the original program sponsors clearly understood something of the implementation problem. FWLERD (Nimnicht, 1968) described a year-long plan that included nine seminars for program assistants—each district was to have one. This proposal described a staff of five persons to develop materials and conduct the seminars and stated a five year minimum for program implementation in a given district. The Universities of Kansas and Pittsburgh described a three stage program including orientation, training institutes for supervisory personnel, and four to six week visiting internships for teachers and/or supervisory personnel at the Pittsburgh and Kansas developmental schools (Resnick, Bolvin, & Bushell memo, February 21, 1968). The Engelmann-Becker group described a three phase teacher training program—(1) a six week intensive summer program for all teachers, (2) field supervision, with a ratio of one supervisor for twelve teachers, and (3) a demonstration class that would pilot the materials for the local projects and that would be visited by local project teachers, in groups of eight, for one week sessions during the first semester. (Bereiter-Engelmann proposal, mimeo, undated) Marie Hughes described a similar approach in a letter to local projects dated March 22, 1968 (Hughes, 1968).

Common to all of these proposals were a central staff, pre-startup training, and continuing training and feedback. Three of the five described a specific, continuing liaison between the sponsor and the local project. Other proposals were reasonably parallel, with some sponsors placing relatively greater emphasis on summer workshops, e.g. Bank Street, and others on a continuing sponsor/project liaison. Two of them, FWLERD and the University of Arizona, explicitly referred to microteaching as one of the techniques to be used.

A reasonable generalization about the Follow Through sponsors might be that although their programs were not as fully developed as research reports and the popular press implied, their developmental capacity and their willingness to work with the tough problems of implementation were promising (Egbert, 1973).

In the summer of 1968 program sponsors began working with from one to fourteen separate projects, with a given sponsor's projects likely to be scattered across the entire country, though the projects of some sponsors tended to be concentrated more in certain parts of the country than did others. More than a dozen institutes and workshops ranging from one to six weeks in length were held during that first summer. Over the next several years, scores of workshops were held, and thousands of visits and telephone calls were made to and from local projects. From this experience, numerous changes were made both in programs and in delivery systems. By the winter of 1970-71 when sponsor reviews were conducted by USOE staff, substantial progress had been made by most sponsors in both of these major dimensions of their responsibilities. Problems identified seemed to be divided fairly evenly between the two (Follow Through Program Review Memoranda, December 1970 to May 1971).

Comprehensive descriptions have been published of Follow Through implementation processes (Krulee, Hetzner, & McHenry, 1972; USOE, The FT Planned Variation Experiment, Vol. III; and Elliott, Judd, & Wood, 1975). An anthropological interpretation was reported by Beers (1976).

Community Selection

In a separate meeting, local, state, and federal education and Office of Economic Opportunity representatives decided that:

1. Communities could be pre-selected to participate in Follow Through, if States were involved in the selection. Both USOE and OEO preferred an alternative in which proposals would have been accepted from all eligible communities, i.e., those having full year Head Start or similar "quality preschool program." Follow Through administrators judged this to be both burdensome and counter productive. The counter productive issue was twofold: (a) The "Tampa" round of pilot selection suggested, that if all local districts could submit proposals, the entire selection process could be politicized, and (b) districts with more resources, i.e., those with more money, tend to submit better proposals, thus suggesting that those districts in greatest need would be less likely to be chosen.
2. Communities could be required to choose from a restricted set of program approaches, associate with a sponsor, and accept the assistance of the sponsor in developing and implementing that approach;
3. Communities could be required to contribute an amount of Title I money equal to 15% of the EOA grant or 10% of the Title I grant, whichever was less; and
4. Communities could be required to involve parents and other community members in program planning and operation.

Although each of these decisions was important to Follow Through and each came to be generally accepted, each was an issue of some concern at that time (Follow through meeting notes, Dec. 11-12, 1967).

After this meeting was completed, the process of identifying potential new Follow Through communities took place. From approximately 225 school districts nominated jointly by state educational agencies and economic opportunity offices and reviewed by regional selection panels, 51 new communities, in addition to the forty 1967-68 pilot projects, were invited to participate in Follow Through's program of planned variation. Preference was given to communities with high concentrations of poverty. Within a given community or district schools with the highest concentrations of graduates from a Head Start, or comparable preschool program, were selected (Estes, March 13, 1968).

Two meetings were held in Kansas City, Missouri, February 20-24 and 25-28, 1968—the first for representatives from the pilot projects and the second for representatives from the prospective sites. The meetings were designed to acquaint participants with the new phase of planned variation and for them to select the approach they wanted to adopt. Pilot projects from 1967-68 were given the option of participating in the new phase or continuing with their original pilot plans. New communities were required to select one of the program approaches (Egbert, 1981).

Program Funding and Changes from 1968 to 1972

In succeeding years additional communities—approximately seventy in 1969 and eighteen in 1970—were brought into the program, following selection procedures paralleling those used in 1968. This program expansion was not sought by Follow

BEST COPY AVAILABLE

Through; it resulted from OEO, HEW, and BOB decisions to expand Follow Through funding as follows: $3.75 million in FY 1967; FY 1968, $11.25 million; FY 1969, $32 million; FY 1970, $70.30 million; and FY 1971, $69 million (Division of Follow Through, undated [1982]). In fact, because Follow Through administrators recognized the management problems associated with increasing the number of local projects as well as the danger that future funding could be reduced, they controlled the increase in the number of Follow Through projects by forward funding some projects. In other words, the natural expansion, given the increased funding would have been greater than seventy projects in 1969. Follow Through prevented this greater expansion.

The initial set of sponsors was expanded to 20 in 1969-70. Besides providing somewhat different approaches to working with children, five of these six new sponsors gave opportunity for representation by three different groups—a state education agency, minority colleges, and a profit making company—not included in the first set of sponsors.

The decision to bring these five sponsors into Follow Through was not entirely programmatic. The USOE has always been anxious to ensure that states have an opportunity to play a major role in federally funded programs. Follow Through provided an opportunity to explore a new state role, that of program sponsor. California was chosen to fill that role. In 1968, too, private businesses sought inclusion in new education ventures. A number of persons in the federal government wanted them to have this opportunity. Follow Through provided a convenient vehicle for giving them such experience. Finally, none among the initial fourteen sponsoring organizations truly represented ethnic minorities. At a meeting of Follow Through general consultants (Each project had a general consultant who helped with parent involvement and other non-sponsor issues.) in Atlanta in October, 1968, Black general consultants caucused. After this caucus and a follow-up meeting of a few Black consultants, Follow Through agreed to locate appropriate minority sponsors, give more attention to community involvement, employ additional minority persons on the staff, and consider minority concerns in the evaluation.

Three primarily Black institutions were selected as Follow Through sponsors in 1968-69, and in 1969-70 developmental grants were made to a fourth Black potential sponsor who proposed a role-trade model for Follow Through and to a Chicano educational psychologist who proposed development of a new approach to bilingual-bicultural education.

Perhaps of equal importance to the agreements reached following Atlanta was the effect of these meetings on Follow Through staff, sponsors, and general consultants. This meeting followed Dr. King's assassination by only six months and feelings were strong and powerfully expressed. For some, this was their first direct contact with Black anger, articulately and personally expressed. Few people left Atlanta unchanged (Egbert, 1981).

By the early 1970s, the administration had established new priorities and sought ways of reducing on-going programs to secure support for its priority programs. Beginning in 1973 attempts were made to reduce Follow Through funding and to phase out the program (Bell, 1974).

Follow Through Evaluation: Phase 2—1968-1972

The Follow Through evaluation has been the subject of numerous reports and critiques, e.g. House, et al. 1978; Elmore, 1975; Haney, 1977; Krulee, et al. 1973; McDaniel, 1975; Cohen, 1975; Mosteller, 1975; Bereiter & Kurland, 1978. Quite correctly, most of these reports have dealt with both the overall program design and with what the government termed the evaluation. Although in some sense the two (the program design and the evaluation) are separable, in other ways the various evaluation efforts were so dependent on program decisions that they are best considered together.

Program Decisions and The Follow Through Evaluation

The most commonly noted problem in the Follow Through program is that it did not meet the requirements for an experiment, e.g. communities were not randomly chosen and assigned to different sponsors, and, within projects, neither the school nor the child was randomly assigned to Follow Through (FT)/non-Follow Through (NFT) conditions. The non-random decision was made explicit in the beginning, and both internal and external evaluation experts either explicitly or implicitly accepted this reality.

There were several reasons for using non-random choice and assignment procedures. Among these were:

1. The procedure of SEA/STA (State Technical Assistance— the OEO office at the state level) nomination that was then followed by negotiated selection between state and federal offices was used for the following reasons.
    a. State offices possessed the essential information about the location and size of Head Start and Title I preschool projects.
    b. EOA required (l) a mixture of rural and urban projects, and (2) non-public school involvement (OEO officials were the interpreters of these requirements. e.g., Davies to Egbert memo, Dec. 9, 1968).
    c. The memorandum of agreement between OEO and USOE required that at least half of the children served in a FT project be graduates of a Head Start or other quality pre-school program. Only through careful tracking and extensive negotiation could these requirements be met.
    d. Both the Tampa round of selecting Follow Through projects and the consistent victories, in other programs, of STAs and SEAs working through their elected representatives on the Hill provided convincing evidence that wherever possible potential problems should be worked through in advance with state agencies.
2. The procedure of permitting the local project to select its own program approach was used for the following reasons.
    a. Community action principles suggested that those involved in the local project should have as much control as possible.
    b. Tradition dictated that local schools control their own curriculum and teaching methods.

c. Follow Through staff felt that if the project made its own model selection it would have a greater commitment to that program approach.

d. Some State and local advisors were uneasy about requiring even that local projects adopt a sponsor. Only after extensive discussion did they concur that, if given a choice, the project could be required to implement a program approach developed outside of the community (Follow Through Meeting notes Dec. 11-12, 1967). (There's little question that, if some of the pilot projects had been required in 1968 to choose a sponsor, those projects would have appealed to powerful political figures. e.g. Portland, Tampa, New York. Had they done so, the entire program could have been scuttled.)

3. Children were not assigned randomly for essentially the same reasons that projects were not assigned randomly to program sponsors. In addition, assigning children randomly would have required extensive inter-school and inter-district assignment and transportation.

4. In the mid-1960s researcher/program developers were reporting remarkable success with new approaches to working with children. It seemed likely that these approaches which had worked so well in controlled experiments would also succeed in Follow Through.

From the first, program administrators stated that because the Follow Through planned variation program was not a traditional experiment special accommodations would be necessary in the evaluation (Egbert to Hughes research and evaluation budget justification memorandum, February 19, 1968).

> *It must be re-emphasized that the "planned variation" design of the Follow Through program by no means approximates the paradigm of the controlled experiment. Hence, it will be particularly important that we obtain measures, not only of the educational and developmental accomplishments of the children and their families, but also of the processes which each community has succeeded in putting into effect. The types of information needed for these assessments...are costly to obtain but essential for inferences and conclusions needed for future Program guidance... (W)e are talking here about such time-consuming procedures as direct classroom observation, observations of small groups of children in special settings and lengthy interviews with school administrators, teachers and parents. These activities require large resources even when the most efficient sampling designs are used.*

This perspective was reinforced in an AERA presentation by Follow Through's Chief of Research and Evaluation (Snyder, 1969, pp. 11-12).

> *The paradigm for the ideal field experiment would call for a situation in which a small number of clearly defined parameters or program elements could be systematically varied, in which self-selection played no part in determining which treatment was adopted by the communities, and in which each experimental group was matched with a true control group—i.e., in which the distinction between experimental and control groups was determined by random assignment. However,*

although inherent limitations prevent even approaching this paradigm in Follow Through, we believe that it will be possible to make useful comparisons among the different approaches; we also believe that the experience of the program sponsors will soon permit them to define much more sharply a number of issues which can be studied with more systematic research designs, either on an intra or inter project basis.

"Action research" always involves compromises; and by now you must see how complicated a task we face in trying to study the development and measure the impact of the different program approaches in ways which will yield the kind of information on which decisions about future program development can be based.

Planning for the Follow Through Evaluation

Serious planning for the Follow Through evaluation began at a March 21-23, 1968 meeting of HEW and OEO staff and such external consultants as David Cohen, Eugene Glass, Robert Hess, Thomas Hastings, Halbert Robinson, Michael Scriven, Susan Stodolski, Robert Thorndike and Edward Zigler. As reported in a long memorandum by one of the participants (Wyatt to Rivlin and Wholey, March 25, 1968) a number of issues were discussed, agreements were reached, and next steps were decided upon. Among the decisions were: (a) There was not enough time to prepare and issue an RFP; contacts would need to be made with a restricted number of organizations inviting them to prepare proposals. (b) Some minimum, common pretest data should be obtained in the fall of 1968. (c) Additional child measurement instruments were needed. (d) (strong agreement) There are two stages in the development of a program approach and it is important to distinguish between them in the evaluation. During the first (formative) stage, which could last for several years, the evaluation should focus on the development process as much as on "outcomes" with children. A comparison between programs, to the extent it is possible (this was not fully resolved), can begin to take place *only after* (Emphasis in original) a period of time, as programs become comparatively less fluid and more summative. (This two-stage concept was explicitly recognized in Follow Through's 1970 "Reporting Plan for Follow Through Evaluation" from Egbert to Hereford, June 24, 1970.)

Of interest also in Wyatt's (March 25, 1968) summary is that at least some suggestion was made that teachers and other project populations, including students, should have some input into the evaluation. No reference is made in Wyatt's summary that anyone objected to non-random assignment of projects, schools, or children.

Following the March 21-23 meeting discussions were held with a number of organizations considered capable of conducting the Follow Through evaluation. Three of the organizations actually submitted proposals—Stanford Research Institute (SRI), American Institutes for Research in the Behavioral Sciences (AIR) and Educational Testing Service (ETS). The AIR proposal was brief, explicit, and specific; the SRI proposal was longer, and contained a more complete discussion of issues and problems; the ETS proposal was brief (four pages), vague, and suggested that the first year should be spent in planning, exploration of centers, public relations, identification of sample, "beginnings of professional cooperation with centers," and, in

BEST COPY AVAILABLE

May, collection of child baseline data. The SRI and AIR proposals were precise documents with the usual institutional capability boilerplate; the ETS proposal was headed "Notes on Possible ETS Participation in Follow Through Research and Evaluation" and was reproduced on a spirit duplicator. The AIR and SRI proposals were explicit about staff and made half time and greater commitments of senior staff members; ETS was vague in its staff commitments (Egbert, 1981).

The review panel, including outside experts as well as representatives of OASPE, OPPE, BESE and RP/E (OEO) agreed unanimously on SRI (Hughes to Estes memorandum, June 25, 1968). Retrospective judgment suggests that the ETS proposal, with an entire year spent on planning, community relations, sample selection, etc., might have been a better choice. But no one recognized that in 1968.

Richard Snyder, Evaluation Section Chief in Follow Through, recognized the need for immediate and sustained help in assessing and further planning of the Follow Through evaluation. He arranged for a contract with the Social Science Research Council (SSRC) to provide this kind of support. The contract to SSRC was a serious effort to provide the continuing, strong conceptual support and guidance recommended by the March 21-23, 1968 panel and suggested by others later on as having been needed. Unfortunately, this contract was not successful in its major objectives. Issues were raised, but resolutions were not achieved. Later on Huron Institute was given a contract to provide specific design assistance.

Although most of the publicity surrounding the Follow Through evaluation has been concentrated on child outcomes as measured by off-the-shelf tests of achievement and non-achievement matters, Follow Through made serious attempts to provide context to the child outcome measures and to extend and improve the measures available. The range of efforts that Follow Through attempted as it struggled to comprehend and master the conceptual, developmental, administrative issues involve in the evaluation has been noted by Haney (1975). These efforts are discussed below under the following headings: (1) Follow Through program implementation; (2) model description and implementation; and (3) child outcome measures.

Follow Through Program and Project Implementation

Studies of overall program and project implementation included (1) overall program organization and implementation (Krulee, 1973; Beers, 1976; Elliott, Judd, & Wood, 1975); (2) case studies of individual Follow Through projects including important interactions with various persons and organizations external to the local project (Stanford Research Institute—Crockett); (3) a review of the health services actually available to children at the project level (Bio-Dynamics—Sullivan); (4) parent interviews with samples of Follow Through and non-Follow Through parents (Stanford Research Institute, March 1971); (5) support systems at the local project level (Institute for Applied Behavioral Science—Fox, March 1971); (6) teacher characteristics and attitudes (Stanford Research Institute); and (7) a cost analysis (RMC).

Model Description and Implementation

Krulee, et al. (1973), Judd and Wood (1973), Elliott, Judd, and Wood (1975), and the sponsor volume of the USOE planned variation experiment report all attempted

to document some portion of sponsor behavior. However, there also were at least three other attempts to achieve some level of sponsor documentation. Two of these were major undertakings at analysis, comparison, and assessing the achievement-related effectiveness of classroom processes. The third was a questionnaire request for sponsors to judge the degree of project and classroom implementation of their models.

Both SRI (Stallings, 1975) and the University of Florida (Soar, 1971; Soar & Soar, 1971) received contracts to study classroom processes. Soar used the same sorts of observation instruments that he had in other studies. Stallings constructed new instruments which were explicitly intended to reflect teacher and pupil behaviors which different sponsors valued and sought to establish. Both Soar and Stallings found differences among the various sponsors; they also found some marked similarities and considerable overlap on many dimensions between classrooms from different sponsors. In addition to their model description and implementation results, both Stallings and Soar found achievement test score relationships with various classroom processes.

One of Stalling's major tasks was to examine the degree of implementation that occurred in each of several projects for each of seven sponsors. To accomplish this task she constructed certain measures that reflected what sponsors did but that also differentiated among sponsors. She then applied the measures to a sample of Follow Through and comparison first and third grade classrooms in each of five projects for each of the seven sponsors. The results of this sub-study are summarized in Table 1 (Stallings & Kaskowitz, 1974).

Data shown in Table 1 indicate that all program sponsors had achieved in excess of 70 percent scores on model implementation by the 1972-73 school year. The University of Kansas was highest with implementation scores above the 80 percent level. Although there are fundamental ways in which elementary classrooms are

Table 1
Implementation Scores for First and Third Grades, by Model Sponsor and for Non-Follow Through Comparison Classrooms

Grade	First					Third			
Sponsor	FT	Non-FT	t=	p<		FT	Non-FT	t=	p<
Far West	78.3	60.3	11.18	.001		76.4	59.0	7.18	.001
Univ. of Arizona	73.6	61.8	11.76	.001		72.3	60.7	4.77	.001
Bank Street	74.8	62.7	7.12	.001		68.5	62.4	3.20	.001
Univ. of Oregon	78.2	61.0	6.11	.001		76.5	60.4	5.62	.001
Univ. of Kansas	84.6	62.4	9.22	.001		83.3	61.3	8.89	.001
High/Scope	76.6	63.7	7.58	.001		75.0	63.5	5.93	.001
Educ. Dev. Cntr (EDC)	76.9	61.2	5.35	.001		75.4	60.7	5.18	.001

BEST COPY AVAILABLE

similar, Stallings and Kaskowitz implementation scores, Table 1, indicate that comparison classrooms also clearly differ from the classrooms of each model sponsor.

Besides examining model sponsor classrooms implementation scores in relation to those of non-Follow Through comparison classrooms, Stallings and Kaskowitz (1974) completed an analysis to see whether classrooms could be classified by sponsor. Of a total of 524 classifications, 410 were correct. University of Kansas and University of Oregon classrooms were rarely identified with another sponsor and classrooms associated with other sponsors were rarely classified as being Kansas or Oregon. Very few non-Follow Through classrooms were classified as University of Kansas classrooms. "They were most often classified as EDC in the first grade and as University of Oregon in the third grade, on the grouping and activity variables. On the interaction variables, the non-Follow Through classrooms were distributed rather evenly across Far West, University of Arizona, Bank Street, University of Oregon, High/Scope, and EDC" (p.341).

Funding of the Soar and Stallings studies began in the first year of the Follow Through evaluation program, 1968. In addition to these two major studies, SRI obtained fairly extensive data from sponsors concerning implementation in both 1969-70 and 1970-71 (Sorenson memo to Egbert & Snyder, April 9, 1971). Sponsor judgments were obtained on the degree of classroom and project implementation; ratings also were obtained for individual teacher performance. In 1969-70 the classroom/project ratings were undifferentiated as to reason for the judgment; however, the 1970-71 evaluation asked for the criteria used in making the judgment and for a weighting of those criteria.

Child Outcomes

Although, as indicated above, Follow Through, as part of the War Against Poverty, was vitally interested in institutional changes that might be expected to affect schools, programs, and families as well as child outcomes, primary attention has been given by others to gains that Follow Through children made, primarily on measures of school achievement. Rivlin and Timpane (1975), for instance, in their assessment of the information-producing effectiveness of the planned variation strategy, wrote of the ideal of identifying "several competing approaches to improving the performance of children from deprived homes"...and then "carefully and methodically set(ting) up the Follow Through program to accomplish these objectives"(p. 3).

The Follow Through evaluation effort was designed to document various sorts of institutional change; it also was intended to assess child outcomes. Among the most perplexing of the issues faced by that part of the evaluation was the nature of the instruments available to measure child characteristics and performance.

Instrument Development

Several program sponsors, as well as Follow Through evaluation staff members, were dissatisfied with instruments available for measuring child characteristics and performance. One problem was with the instruments themselves, but an equally

important issue was in deciding what characteristics were important to measure. Some sponsors wanted to focus on achievement outcomes and were generally satisfied with available achievement tests; other sponsors wanted to add to the range of characteristics measured and were dissatisfied with the available tests; still other sponsors were opposed to measuring child outcomes. In fact, some sponsors opposed any sort of external evaluation—and, in some instances, this included their own external evaluation of the children and projects with which they were working.

A number of sponsors had constructed instruments that they used in their own programs, and in some instances they wanted progress in their projects to be judged by improvements in scores on these instruments. This led to one of the early decisions made in the evaluation which was that all children in the sample would take a certain set of tests and that there also would be sponsor-specific tests that would be used only in the projects working with a specific sponsor (Emrick, Sorenson, & Stearns, 1973). This decision created considerable uneasiness among administrators responsible for Follow Through.

Two sources of serious dissatisfaction with the tests available were with those designed to measure (1) complex cognitive processes and (2) non-cognitive (e.g., social and personal) processes. A grant was given to Educational Testing Service to identify important complex cognitive processes and develop one or more instruments intended to measure those processes; funds were added to the basic Stanford Research Institute contract to develop one or more instruments for identifying and assessing non-cognitive processes. A grant also was given to Columbia University to develop measures of important child characteristics from observing the child in the classroom. None of these efforts produced the desired instruments. Although each of these undertakings was more or less successful in further defining the nature of the task, each one failed in its task of producing an instrument useful in the Follow Through evaluation (Egbert, 1981).

Follow Through Program: Phase 3—1972-1982

Follow Through's peak program years were in the decade of the 1970s —from FY 1970 to FY 1979. Follow Through began the decade with a FY 1970 appropriation of $70.3 million and ended it with a FY 1979 appropriation of $59.0 million. These figures, along with others, are shown in Table 2 (Follow Through, undated [1982]).

Of importance also in Table 2 is the contrast between the funds available and the number of children for which local projects were funded between 1968 and 1972. Funds available almost tripled between school years 1968-69 and 1969-70 yet the number of children who were funded little more than doubled. Funds available for school year 1970-71 were almost 2½ times as great as for 1969-70, yet the children funded were little more than 1½ times as great. On the other hand, the number of children funded continued to increase through the 1972-73 school year even though funds were down 10 percent from the peak reached in 1970-71 school year. This phenomenon resulted from the decision of Follow Through administrators to fund as many projects as possible for two years in order to keep the program from growing more rapidly than necessary.

BEST COPY AVAILABLE

Table 2
Appropriations, Number of Sponsors, Number of Local Projects, and Number of Low Income Children Funded–Follow Through (FY 1968-1982)

Fiscal Year	Approp. in Ms.	School Year	No. of Sponsors	No. Local Projects	No. of Low-Income Children	Grades of FT Classes
1967	3.75	1967-68	0	39	2,900	K-1
1968	1.25	1968-69	14	92	15,500	K-1
1969	32.00	1969-70	20	160	37,000	K-2
1970	70.30	1970-71	22	178	60,200	K-2
1971	69.00	1971-72	22	178	78,170	K-3
1972	63.06	1972-73	22	173	84,000	K-3
1973	57.70	1973-74	22	170	81,000	K-3
1974	53.00	1974-75	22	169	78,000	K-3
1975	55.50	1975-76	22	165	76,500	K-3
1976	59.00	1976-77	20	164	75,700	K-3
1977	59.00	1977-78	20	161	74,675	K-3
1978	59.00	1978-79	19	157	70,500	K-3
1979	59.00	1979-80	19	153	68,819	K-3
1980	44.25	1980-81	19	147	63,558	K-3
1981	26.25	1981-82	16	84	36,319	K-3
1982	19.44	1982-83	15	69	N/A	K-3

Follow Through Evaluation: Phase 3—1972-1977

The Stanford Research Institute held the entire, major Follow Through Evaluation contract through 1972. From 1972 on, however, SRI's role became restricted largely to data gathering. Abt Associates secured the data analysis portion of the contract in 1972 and completed the final set of reports in 1977.

The activities, successes and failures, and results obtained from the Follow Through evaluation, largely through the SRI and Abt contracts, have been described, analyzed, criticized, defended, and commended, e.g. House, et al., 1978; Anderson, 1977; Anderson, et al., 1978; Wisler, et al. 1978; Hodges, 1978; Hodges and Shehan, 1978; Elmore, 1975; Haney, 1977; Cohen, 1975; Mosteller, 1975; Kennedy, 1978; and Bereiter and Kurland, 1978.

The Abt Associates Report

As Mary Kennedy (1978) has said, the volumes of the extraordinarily detailed and complex Abt report make a stack almost six inches high. Volume IV-A, National Evaluation: Patterns of Effects, describes the background of the program and the

BEST COPY AVAILABLE

evaluation (Chapter 1), educational strategies and measures of effectiveness (Chapter 2), the sample (Chapter 3), the analysis strategy (Chapter 4), and patterns of effects (Chapter 5).

In Chapter 5, Volume IV-A, of the Abt report, Stebbins (Stebbins, et al., 1977) presents and describes ten findings.

1. The effectiveness of each Follow Through model varied substantially from site group to site group; overall model averages varied little in comparison.
2. Models that emphasize basic skills succeeded better than other models in helping children gain these skills.
3. Where models have put their primary emphasis elsewhere than on basic skills, the children they served have tended to score lower on tests of these skills than they would have done without Follow Through.
4. No type of model was notably more successful than the others in raising scores on cognitive conceptual skills.
5. Models that emphasize basic skills produced better results on tests of self-concept than did other models.
6. Model comparisons in New York and Philadelphia yield results which are similar to those found in overall comparisons.
7. Some models are more successful in their most disadvantaged sites.
8. Two models are consistently more effective with Head Start children.
9. Most models are more effective during kindergarten and the first grade than during second and third grade.
10. Some Follow Through sponsors grew in effectiveness over time (p. iii).

Essentially, then, the Abt report message was that the measured outcomes variation among sites within program models was greater than the variation on those same outcomes among models. Within this general finding, program models that emphasized basic skills produced better results on tests of (1) basic skills, and (2) self-concept than did other models, and children in models that did not emphasize basic skills tended to score less well on tests of those skills than they would have done without Follow Through. No type of model was more successful than other models in raising scores on cognitive conceptual tests.

A special concern of Follow Through was with those children who were most disadvantaged. Thus, the Abt finding that some models were relatively more effective with these children was especially important. Direct Instruction, Parent Education, Behavior Analysis, Bank Street, and EDC all had their greatest impact on the most disadvantaged children on both basic skills and cognitive conceptual achievement domains as well as on affective measures. In addition, Responsive Education had its greatest impact on the most disadvantaged children on the basic skills and cognitive conceptual domains; California Process had its greatest impact on the most disadvantaged children on affective measures.

The findings from comparison of results with Cohort II Kindergarten and Cohort III Kindergarten in general seem to indicate that the length of time that a model was

implemented in a site was positively related to the model's effectiveness as measured by scores on basic skills and cognitive conceptual instruments. Within this general finding, however, is the indication that those models where children in Cohort II-K performed least well were able to show greatest gains for children in Cohort III-K.

The House Critique and the Abt and Federal Government Rejoinders

Apart from the Abt report itself, particularly Volume IV-A (Stebbins, St. Pierre, Proper, Anderson & Cerva, 1977), perhaps the best known of the reports and comments about the Follow Through evaluation appeared in the *Harvard Education Review* , Volume 48, Number 2—1978. This issue was devoted entirely to the House, et al. critique of the Follow Through evaluation and to three responses to that critique.

The House (House, et al., 1978) critique of the Follow Through evaluation was especially negative, and it led to hostile rejoinders by Abt (Anderson, St. Pierre, Proper, & Stebbins, 1978). House commented negatively on everything from the Follow Through program design (large numbers of sponsors, sites and children and no systematic selection of schools within sites or teachers within schools) to the report's use of the word "basic" as in basic skills. Stebbins (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977) applied "basic skills" primarily to those parts of the curriculum that often times are taught through rote methods. House (House, et al., 1978) commented that labeling such skills as basic "gives them an importance in relation to the others that they do not deserve" (p. 137). House was especially critical of the Abt report's model classification scheme, including the classification definitions, the tests used to measure outcomes (They were too narrow.), the "flawed" statistical analysis (primarily analysis of covariance), and the confounding of program effect with size of groups in some of the analyses.

House criticized, too, the program's failure to be sensitive to the concerns of parents and sponsors, but he also faulted the sponsors. "Although treated unfairly, the sponsors were not blameless. In spite of dissatisfaction with the evaluation, they continued to . . . receive large sums of money from the government" (p. 132).

Writing for Abt, Anderson (Anderson, St. Pierre, Proper, & Stebbins) responded in kind to the House critique and refuted each criticism in turn. At one point Anderson compared the House critique to the newspaper column that appeared to him to be the point of departure used by House. "We provided the House panel with a complete set of our report documents, and we spent a good deal of time helping its chairman understand the context of our findings. Since our colleagues have chosen to adopt Kilpatrick's strange distortion of emphasis, we can only conclude that they are really criticizing *his* findings, not ours. If that is indeed what they are doing, we would have been grateful if they had done it much more explicitly" (Anderson, et al., 1978, p. 164).

Abt (Stebbins, et al., 1977; Anderson, et al., 1978) and House (House, et al., 1978) agreed on two major points, though these points sometimes appeared to be lost in the heat of the exchange. First, the measured outcomes variation among sites within program models was greater than the variation on those same outcomes among models, and second, in most cases, the Follow Through groups scored about as one would expect similar groups of children to score on the instruments

used without Follow Through. In other words, the evaluation produced no convincing evidence that the Follow Through strategy is an effective approach for raising poor children's scores on these instruments.

When Haney (1977) noted, "...if the (Follow Through) evaluation has improved over time, it has been getting better at answering narrower and narrower questions." (p. 249), he was not commenting on Abt's role in the evaluation; nevertheless, he did address the dilemma that Follow Through faced. It was the dilemma of a program that was too large and complex to "evaluate" with a single, simple design and set of measures, yet the times demanded that such an evaluation be attempted.

Wisler (Wisler, Burns, & Iwamoto, 1978) presented the government's view of the evaluation. In essence, Wisler acknowledged that problems existed in the program design and made the case for more highly controlled studies, but he also characterized the House critique as being misleading in what the evaluation findings were and what conclusions could be drawn from them. Where House said that the cognitive instruments were too narrow and should have been augmented, Wisler argued that testing time already was at a maximum, especially for comparison groups. Where House said that the affective instruments were not valid, Wisler described the process used to select them and said that they were the best available. In short, Wisler, like Anderson, refuted the House panel criticisms point by point.

Other Comments and Re-analyses

Two reports that commented on the Follow Through evaluation and then reported re-analysis of the data are quite useful (Kennedy, 1978; Bereiter & Kurland, 1978). Kennedy performed a meta-analysis of the Abt data. Her findings paralleled certain key Abt findings, e.g. most models tended not to show outcome measure differences between Follow Through and non-Follow Through children, and the two models that had at least one large positive effect were both structured classroom approaches.

In addition to giving findings, Kennedy suggested alternative explanations for finding positive and negative effects and then presented arguments, and sometimes data, to support her explanations. One potential explanation for negative effects for a given sponsor, for example, was that the instruments used did not measure what the sponsor was trying to accomplish in a given area. In support of this explanation, Kennedy reported both Abt (Metropolitan Achievement Test [MAT]) and sponsor (Cognitive Curriculum) data for language achievement. The MAT focused on grammatical correctness; the sponsor measure focused on children's communication power, e.g. number of words used in a writing sample, proportion of descriptive words used in the writing sample. The results when the sponsor instrument was used were much more positive for Follow Through children.

Although some critics (e.g. House, et al., 1978; Mosteller, 1975) have questioned the validity of any results derived from the Follow Through/non-Follow Through population, Bereiter & Kurland (1978) have argued quite convincingly that this perspective is unduly pessimistic. Their re-analysis of the Abt data shows a clear "victory" for the behaviorally oriented programs. As an argument for further studies of the Follow Through child population, the important issue of the Bereiter and Kurland study is not "who won" but that there is enough potential power to permit finding inter-sponsor differences. (Bereiter & Kurland finessed the FT/NFT issue in

BEST COPY AVAILABLE

their analysis; however, as the Abt report is viewed by at least some observers, e.g. Wisler, et al., 1978, it appears to provide similar justification for comparing FT/NFT groups.)

Followup Studies

The Abt report of student outcomes at the conclusion of the Follow Through experience was important; conducting followup studies sometime after program completion also is potentially useful. At least five sponsors and persons associated with sponsors (Arizona, Oregon, Southwest Education Development Laboratory [SEDL], Bank Street, and University of Florida [later, North Carolina]) have completed a limited number of followup studies of Follow Through graduates and comparison groups; Philadelphia, which worked with six different sponsors, has published several followup studies; and two summaries (Olmsted & Szegda, 1987; Wang & Walberg, 1989) of followup studies have been reported.

All of the followup reports have employed one or more measures of (1) grade level retention or special education assignment, (2) attendance or dropout, and (3) school achievement. In general, the reports have shown positive results for Follow Through graduates on each of these three general dimensions. Thus, in the studies Follow Through graduates have had significantly fewer retentions at grade level or a lower percent of special education assignments (Cloud, Rentfrow, & Hildebrandt, 1979—University of Arizona; Gersten, Carnine & Keating, 1987—University of Oregon); significantly higher percent of attendance or lower dropout rates (Szegda, 1986—University of Florida [North Carolina]; Gersten, Carnine & Keating—1987 University of Oregon; and Philadelphia, 1982); and significantly higher scores on various achievement measures (Szegda, 1986—University of Florida [North Carolina]; Gersten, Carnine, & Keating, 1987—University of Oregon; and Philadelphia, 1982).

Two of the studies (Cloud, Rentfrow, & Hildebrandt, 1979—University of Arizona; Szegda, 1986—University of Florida [North Carolina]) used older siblings as the comparison groups; the others all used children of comparable socioeconomic status from the same grade levels and usually from the same community.

Fragmentary though these followup studies of Follow Through graduates are, they suggest that further studies, systematically planned and conducted, could provide a mosaic of useful information. They also offer guidance for selection of alternative comparison groups as well as suggestions for social and behavioral indices of program outcomes.

Follow Through Program: Phase 4—1981-1991

In February of 1981, the National Institute of Education held a conference in Philadelphia to "provide input to assist NIE in planning a new program of Follow Through research and development" (Kocher, January 5, 1981). Papers were prepared and presented and extensive discussions were held during this two day period. Based on these discussions, additional sponsors were brought into the Follow Through program.

A later meeting was held in September of 1982 (Stalford, September 10, 1982), to discuss the possibility of longitudinal evaluation of Follow Through. This meeting did not lead to such an evaluation.

Discussion

From the very first meetings that led to Follow Through's planned variation approach and the use of program\sponsors, some of those involved have argued that anlysis based on data gathered at the end of third grade would be inconclusive at best and that it could lead to conclusions the opposite of those which should be reached.

The arguments against using third grade data as the ultimate "test" are three-fold.

1. Gains resulting from the Follow Through experience might wash out within a year or two after the program's conclusion.
2. The success of education, particularly Follow Through, cannot be measured by test scores alone. Other school and out-of-school indices are equally as important as test scores. These other indices, e.g. school attendance and graduation, delinquency, employment, etc., do not have readily identifiable correlates at the third grade level.
3. The third grade may not be a good time to take measures even on traditional achievement instruments. Although there is a correlation between achievement test performance at the end of third grade and later scores, there is little evidence of performance on different curriculum models and later performance. However, it seems possible that there would be differences between those models that emphasize content and those that emphasize the processes of learning and development. If this argument is valid, only when the children in a given Follow Through cohort are older would it be possible to judge future achievement. At that time, perhaps achievement tests should be supplemented by instruments designed to measure developmental progress as well.

References

Anderson, R. B. (1977). *The Effectiveness of Follow Through: What have we learned?* Paper presented at the annual meeting of the American Educational Research Association, New York.

Anderson, R. B., St. Pierre, R. G., Proper, E. C., & Stebbins, L. B.(1978). Pardon us, but what was that question again? A response to the critique of the Follow Through evaluation. *Harvard Educational Review.* 48(2), 128-160.

Beers, C. D. (April 1976). *Principles of implementation: The lessons of Follow Through.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Bell, T. H. (September 13, 1974). Memorandum to Assistant Secretary for Education re. Follow Through continuation.

BEST COPY AVAILABLE

Bereiter-Engelmann Participation in Project Follow Through, mimeo, a proposal, (undated).

Bereiter, C., & Engelmann, S. (1966). *Teaching disadvantaged children in the preschool.* New York, NY: Prentice Hall.

Bereiter, C., & Kurland, M. (April 1976). *Were some Follow Through models more effective than others?* Paper presented at the American Educational Research Association, San Francisco.

Boikes, O. (May 21, 1968). Memorandum from Olga Boikes, Office of Economic Opportunity to Robert Egbert, Director, Follow Through.

Bushell, D. (undated). Untitled proposal for participation in Project Follow Through. mimeo.

Cloud, K., Rentfrow, & Hildebrandt. (July 1979). *Later success of Follow Through graduates in Lakewood New Jersey public schools.* Tucson, AZ: Arizona Center for Educational Research and Development.

Cohen, D. (1975). The value of social experiments. In A. M. Rivlin & P. M. Timpane (eds.) *Planned variation in education: Should ee give up or try harder?* Washington, DC: The Brookings Institution.

Coleman, J. (1975). Comment on "The value of social experiments," pp. 173-175 In A. M. Rivlin & P. M.Timpane (eds.) *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.

Davens, E. (June 19, 1968). Letter from E. Davens, M.D., Acting Chairman, National Planning Committee, Project Head Start to Roy M. Fisher, Editor, Chicago News.

Davies, M. (December 9, 1968). Memorandum to Robert Egbert, Director, Follow Through.

Division of Follow Through, undated (1982).

Egbert, R. L. (December 14, 1967). Letter to Gordon Klopf, Bank Street College of Education.

Egbert, R. L. (January 6, 1968). Memorandum to R. Snyder, Head of Planning and Evaluation for Follow Through, concerning nature of sponsor responsibilities.

Egbert, R. L. (January 5-6, 1968). Notes, program model development meeting.

Egbert, R. L. (February 19, 1968). Memorandum to John F. Hughes, Director, Division of Compensatory Education. Justification of $2,150,000 Research and Evaluation Budget from FY 1968 Follow Through Funds)

Egbert, R. L. (October 14, 1968). Memorandum to L. Wyatt and R. Snyder.

Egbert, R. L. (January 1970). New directions in Follow Through: A position paper. Internal document.

Egbert, R. L. to Hereford, K. (June 24, 1970) Reporting Plan for Follow Through Evaluation. (Mimeographed)

Egbert, R. L. (May 14, 1970). Remarks to SRI Review Panel.

Egbert, R. L. (1971). Internal memo to John Evans, USOE.

Egbert, R. L. (1973). Planned variation in Follow Through. (Mimeographed)

Egbert, R. L. (1981). Some thoughts about Follow Through thirteen years later. Paper prepared for NIE 1981 Philadelphia conference about Follow Through. (Mimeographed)

Elliott, D., Judd, D., & Wood, S. (1985, April). *A description of Follow Through sponsor implementation processes.* (Final Report, Contract No. SBC208(a)-73-C-107) to U.S. Office of Education, DHEW).

Elmore, R. (1975). Design of the Follow Through experiment. In A. M. Rivlin & P.M. Timpane (Eds.) *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.

Elmore, R. (June 9, 1972). The politics and administration of an educational experiment: The case of Follow Through. Mimeographed draft.

Emrick, J., Sorenson, P., & Stearns, M. (February 1973). *Interim evaluation of the national Follow Through program 1969-1971.* Menlo Park, CA Stanford Research Institute.

Estes, N. (March 13, 1968). Memorandum from Nolan Estes, Associate Commissioner for Elementary and Secondary Education, to H. Howe, United States Commissioner of Education.

Evans, J. W. (June 19, 1969). Memorandum from John Evans, Chief, Evaluation Division, Office of Economic Opportunity to R. L. Egbert, Director, Follow Through.

Evans, J. W. (April 18, 1974). *Evaluating education programs—Are we getting anywhere?* Invitational Address before the American Educational Research Association.

Follow Through Conference Notes (September 16-17, 1968). Notes on Follow Through long range planning meeting.

Follow Through Meeting Notes. (January 5/6 and January 26/27, 1968). Notes about meetings with potential program sponsors.

Follow Through Meeting Notes. (December 11/12, 1967). Notes on a meeting of local, state, and federal education and OEO representatives to discuss processes for selecting sites for Follow Through projects for projects selecting of program sponsors..

Follow Through Program Review memoranda, December1970 to May 1971.

Fox, R. (March 1971). *An examination of support systems in selected Follow Through schools.* Report from Grant No. 0-9-526610-4632 (100). Washington, DC: Institute for Applied Behavioral Science.

General Accounting Office. (May, 1973). Report to the Congress: Evaluation of the Office of Economic Opportunity's performance contracting experiment.

Gordon, I. J., Olmsted, P. P., Rubin, R. I., & True, J. H. (1978). *Assistance to Local Follow Through Programs* (Annual report to the Office of Education, Grant No. G00-77-01691, 1977-78). Chapel Hill: University of North Carolina, School of Education.

Greenberg, P. (1969). *The devil has slippery shoes.* New York, NY: Macmillan.

Gray, S. W., & Klaus, R. A. (1965). An experimental preschool program for culturally deprived children. *Child Development,* 36 (4), 887-898.

Haney, W. (1977a). The Follow Through experiment: Summary of an analysis of major evaluation reports. *Curriculum Inquiry,* 7, (3), 227-257.

Haney, W. (1977). The Follow Through planned variation experiment. *(Volume V: A technical history of the national Follow Through evaluation).* Cambridge, MA: The Huron Institute.

Herzman, C. (August 3, 1967). Memorandum to the Files on Summary of Meeting on Follow Through from Carol Herzman, Program Officer, Division of Compensatory Education.

Hodges, W. L. (1978). The worth of the Follow Through experience. *Harvard Educational Review.* 48(2), 186-192.

Hodges, W. L., & Sheehan, R. (1978). Follow Through as ten years of experimentation: What have we learned? *Young Children,* 34, 4-18.

Hodges, W. L., Sheehan, R., & Carter, H. (1979). Educational intervention: The role of Follow Through sponsors. *Phi Delta Kappan,* 60, 666-669.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. (1978). No simple answer: Critique of the "Follow Through" evaluation. *Harvard Educational Review,* 48, 128-160.

Howe, H. (July 18, 1967). Letter from H. Howe, Commissioner, U.S. Office of Education to Shriver, S., Director, Office of Economic Opportunity.

Hughes, J. F. (October 18, 1967). Memorandum to Nolan Estes, Associate Commissioner for Elementary and Secondary Education.

Hughes, J. F. (December 29, 1967). *Budget Requests for the Follow Through Program, Fiscal Year 1968 and Fiscal Year 1969.* (Division of Compensatory Education—Unpublished).

Hughes, J. F. (June 25, 1968). Memorandum from John F. Hughes, Director, DCE to Nolan Estes, Associate Commissioner for Elementary and Secondary Education requesting approval of the Stanford Research Institute. *Proposal for A Longitudinal Evaluation.*

Hughes, M. (March 22, 1968). Letter from Marie Hughes, University of Arizona, to Katherine Brumer, Baltimore Public Schools.

Johnson, L. B. (1967). State of the Union Address *Congressional Record* Vol. 113, Part 1 90th Congress 1st session, p. 37.

Johnson, L. B. (1967). Education and Welfare—Message from the President, Congressional Record, Vol. 113, part 3, 90th Congress, 1st session, p. 2882.

Judd, D., & Wood, S. (June 1973). Follow Through materials review. Portland, OR: Nero Associates.

Kennedy, M. M. (1978). Findings from the Follow Through planned variation study. *Educational Researcher*, 7, 3-11.

Kirschner Associates, Inc. (1970). *A national survey of the impacts of Head Start centers on community institutions*. Project Head Start, Department of Health, Education and Welfare (Department of Health and Human Services). Washington, DC.

Kocher, T., & Stalford, C. (December 31, 1980). Letter to Strand I Paper Writers. National Institute of Education. Washington, DC.

Krulee, G. K., Hetzner, W. A., & McHenry, E. J. (1973). *An organizational analysis of Project Follow Through*. Final Report. Evanston, Illinois: Northwestern University.

LaMoure, L. J. (April 18, 1969) Memorandum from L.J. LaMoure Evaluation Operations and Technical Support Services to Karl Hereford, Director, Program Planning and Evaluation, re. Follow Through Guidelines and Application Forms.

Levin, T. (January, 1967). The Child Development Group of Mississippi: A hot section of the quiet front in the war on poverty. *American Journal of Orthopsychiatry*. 37(1), 139-145.

McDaniel, G. (1975). Evaluation problems in Follow Through. In A. M. Rivlin & P. M. Timpane (eds.) *Planned Variation in Education: Should We Give Up or Try Harder?* Washington, DC: The Brookings Institution.

McDavid, J. W., Gordon, E. W., Grotberg, E. H., & Datta, L. (1967). *Project Head Start: Two years of evaluative research, 1965-1967*. Division of Research and Evaluation, Project Head Start. (Mimeographed)

Madow, W., & Sorenson, P. (August 15, 1968). Memorandum to sponsors of Follow Through Programs.

Mosteller, F. (1975). Comment on the value of social experiments. In A. M. Rivlin & P. M. Timpane (eds.) *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.

National Follow Through Budget Presentation, FY 1972. (August 1970). (Mimeographed)

Newsweek, January 29, 1968, pp. 47-48.

Nimnicht, G. (March 2, 1968). The Proposed Model Follow Through Program. Draft. (Mimeographed)

Notes on SRI/Follow Through discussions, August 20, 1969, February 19, 1970, and March 11, 1971.

Office of Economic Opportunity (February 1, 1972). *An experiment in performance contracting: Summary of preliminary results*. Washington, DC: Office of Economic Opportunity.

Olmsted, P., & Szegda, M. (November 1987). Long term effects of Follow Through participation. Ypsilanti, MI: High/Scope Educational Research Foundation. (Mimeographed)

Orton, R. (September 30, 1968). R. Orton, Head Start Director memorandum to R. Egbert, Follow Through Director.

Pines, M. (1966). *Revolution in Learning: The Years from Birth to Six.* New York: Harper & Row.

P.L. 88-452. Economic Opportunity Act.

P.L. 90-222: Economic Opportunity Act Amendments of 1967. 90th Congress, S. 2388, December 23, 1967.

Resnick, L., Bolvin, J. & Bushell, D. (February 21, 1968). Memorandum to R. Snyder & R. Egbert.

Rhine, W. R. (Ed.) (1981). *Making schools more effective: New directions from Follow Through.* New York: Academic Press.

Rivlin, A. M., & Timpane, P. M. (1975). Planned Variation in Education: An Assessment. In *planned variation in education: Should we give up or try harder?* Alice M. Rivlin & P. Michael Timpane, (eds.) Washington, DC: The Brookings Institution.

Rosenshine, B. (1976). Recent research in teacher behaviors and student achievement. *Journal of Teacher Education,* 27(1), 61-65.

Ryan, S. (1974). *A Report on Longitudinal Evaluations of Preschool Programs. Volume 1: Longitudinal Evaluations.* Washington, DC: DHEW Publication No. (OHD) 74-24.

Schiller, J., Stalford, C., Rudner, L., Kocher, T., & Lesnick, H. (October 1, 1980). *Plans for Follow Through Research and Development.* Washington, DC: NIE.

Schweinhart, L. J., & Weikart, D. P. (1980). *Young children grow up: The effects of the Perry Preschool Program on youths through Age 15.* Monograph #7. Ypsilanti, MI: High/Scope Educational Research Foundation.

Shriver, S. (July 13, 1967). Letter from S. Shriver, Director, Office of Economic Opportunity, to Howe, H., Commissioner, U. S. Office of Education.

Smith., M. S. (1975). Design strategies for Experimental Studies. In A. M. Rivlin & P. M. Timpane (eds.) *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.

Snyder, R. F. (1969). Follow Through: Prospects and problems — an overview of some basic issues. An address given at the American Educational Research Association in Los Angeles.

Soar, R. S. (1973). *Final report: Follow Through classroom process measurement and pupil growth (1970-71).* Gainesville, Florida: University of Florida College of Education.

Soar, R., & Soar, R. M. (1971). Empirical analysis of selected Follow Through programs: An example of a process approach to evaluation. *National social studies education yearbook.* Part 2: 229-59.

Sorensen, P., (SRI) (April 9, 1971). Memorandum to Egbert, R. and Snyder, R.

Stalford, C. (September 10, 1982). Letter to Egbert, R. Univ. of Nebraska. National Institute of Education. Washington, DC.

Stallings, J. A. (1973). *Follow Through program classroom observation evaluation, 1971-72.* Menlo Park, CA: Stanford Research Institute.

Stallings, J. A. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. *Monographs of the Society for Research in Child Development.* 40(7-8).

Stallings, J., & Kaskowitz, D. (1974). *Follow Through classroom observation evaluation 1972-73.* Menlo Park, CA: Stanford Research Institute

Stanford Research Institute International. (March 1971). *Longitudinal evaluation of selected features of the National Follow Through Program. Appendix B: Parent Interview Survey, 1969-70.* Menlo Park, CA: SRI International.

Stanford Research Institute International. (1978). *Summary of Follow Through data.* Menlo Park, CA: SRI International.

U. S. Office of Education. *The Follow Through Planned Variation Experiment. (Volume III: Part A-Model Sponsors' Perspective on Follow Through; Part B-Model Sponsors' Chapters).* Washington, DC: U. S. G. P. O.

U. S. 90th Congress, 1st Session. (February 1967). *Congressional Record.* Vol. 113, pt. 2, p. 2882.

Weber, C. U., Foster, P. W., & Weikart, D. P. (1978). *An Economic Analysis of the Ypsilanti Perry Preschool Project.* Monograph # 5. Ypsilanti, MI: High Scope Educational Research Foundation.

Wholey, J. (December, 1979). New Directions for Follow Through. (Mimeographed)

Wisler, C. E., Burns, G. P., Jr., & Iwamoto, D. (1978). Follow Through redux: A response to the critique by House, Glass, McLean, and Walker. *Harvard Educational Review,* 48, (2), 171-185.

Wolff, M., & Stein, A. (1966). *Study I: Six months later, a comparison of children who had Head Start summer 1965, with their classmates in kindergarten (a case study of kindergartens in four public elementary schools, New York City).* Paper presented at the meeting of the American Psychological Association.

Wolff, M., & Stein, A. (1967). Head Start six months later. *Phi Delta Kappan,* 48, 349-350.

Wyatt, L. (March 25, 1968). Memorandum to Rivlin, A. and Wholey, J.

BEST COPY AVAILABLE

Follow Through Graduates Today

Garry L. McDaniels
Skills Bank Corporation

The purpose of this paper is to discuss the value of conducting a follow-up study of students in the Follow Through Program.[1] Instituted in 1967, the Follow Through program now has many graduates who have entered adulthood. Follow Through still functions today but at a greatly reduced capacity compared to the initial year of the projects. This article focuses on the value of studying the graduates of early efforts.

Why is the Follow Through Program a Study Opportunity?

The Follow Through Program provides a unique study opportunity. No other program of its type was ever introduced into the American school system. The program's characteristics supporting this assertion are follows:

[1] In writing this paper, the author assumes a thorough understanding of the Follow Through Program on the part of the reader. The author also assumes that the considerable technical difficulties of longitudinal studies will be detailed by others. Finally, the author assumes that, while these technical problems will be considered, the decision of whether to institute a study will be based more on the expected usefulness of the findings.

1. The program (or "treatment") was comprehensive, well-financed, and aimed at making significant improvements in the behavior of children, families, and schools.
2. The program created a unique and productive marriage between higher education and the local school community.
3. The program's capacity to create change was tested at urban and rural sites, widely distributed throughout the United States.
4. The program's participants are becoming young adults. This creates an excellent time to see if Follow Through met its goal of improving the life chances of these youth.

Each of these characteristics is discussed more fully below.

Follow Through was a Very Powerful "Treatment"

Education experiments are usually characterized by their weaknesses. They are usually directed at a very small part of a child's school day or a very minor part of a child's school life. The Follow Through Program, in contrast, was a massive treatment. It was a comprehensive, multi-year treatment and was adequately financed.

- Follow Through funds were spent on educational and other needed services. Local Follow Through personnel became service coordinators for medical, dental, and social service providers.
- The program was designed to cover three grade levels, but some communities extended the program to cover the entire six or seven years of the elementary school.
- The federal investment was significant. Seven hundred dollars per child was added to the schools' existing commitment.
- Follow Through empowered its participants. All participants were a part of the decision making and service delivery process.

Coordination of Services was a Major Goal

At the local level, Follow Through brought the community's resources together. The Follow Through Program was coordinator of local services for the participating families. Follow Through sites became centers of service coordination years before the fragmented nature of human services became a commonly discussed topic. Students and their families were not expected to sort their own way through the maze of disparate social service providers. Instead, Follow Through program directors entered into coordinating agreements, both formal and informal, to ensure that medical, dental, legal, and other needed services were available.

Utilization of Expert Knowledge at the Local Building Level

One of Follow Through's significant contributions was that it demonstrated a way to bridge the gap between university-based personnel (or for trainers from non-profit educational laboratories) and the community. These change agents were call program "sponsors."

Follow Through sponsors were selected by communities to serve as expert resources in teacher training, community organizing, and almost anything else needed locally. The federal leadership gave the school districts freedom to choose their own sponsors, and this freedom was vigorously accepted. "Hundreds of school district representatives, Head Start Programs, Community Action Agencies, parent groups, and state agencies . . ." descended on Washington to participate in the program (Egbert, 1973).

The list of participants who managed components of the Follow Through Program is a virtual "Who's Who" in education, child development, sociology, psychology, and evaluation. Perhaps most remarkable was how these resourceful experts were able to transfer their philosophies and ideas to such physically distant sites.

These experts brought ideas, but, equally importantly, they participated as partners in the communities. This was probably made possible by the strength and vision of the Washington leadership. "All of the ranking members of the administrative staff were persons with behavioral or social science training." (Rhine, 1973)

Empowerment

At the local level, Follow Through opened up the school system to the community. Parents, grandparents, and community leaders were invited to direct or advise their local Follow Through Programs. This kind of public access to educational decision making was new to many of the communities that accepted Follow Through grants.

The impact of empowering the participants is well-documented in testimonials, case studies, and surveys (Barnes, 1973). Parents and other community members assumed leadership positions in the programs, often as advisory board members or employees. Many parents went back to school or enrolled in training programs. They also became more skilled at being parents. The information and services provided in the program helped them to become more confident parents and more competent teachers of their children (Haney, 1977; Cline, 1974).

Did Follow Through Make a Difference?

There are always lingering doubts as to whether any treatment or intervention is enough; whether the time and money invested was really enough to make a difference. Whether or not Follow represents a sufficient investment depends on some assumptions about how we help each other in this society.

We live in a society where the dominant assumption is that people have to find the energy within themselves to be successful. Our custodial role for one another is quite limited. We believe in helping each other, not taking care of each other.

BEST COPY AVAILABLE

No intervention program can be effective if we assume that the program has to take care of the participants. We can be successful only if we assume that the participants themselves have resources to offer.

The people for whom the Follow Through Program was intended did indeed bring their own resources to the table. Follow Through children, parents, and community participants brought a sense of hope and pride, a willingness to take charge of their own lives, and an abundance of energy. (Barnes, 1973).

Follow Through was designed to support these positive attributes. It allowed people to make decisions and take control. It enabled them to provide their children with a positive, comprehensive school environment staffed by competent, caring adults.

Follow Through and Head Start were a part of these families' lives for a number of years. Case studies created in the 70s show that lives of families took different direction as a result of these empowerment programs. Follow Through's young adults can show us how useful a comprehensive, expertly organized, empowerment program led by educational institutions might be.

Follow Through Characteristics and Longitudinal Research

Follow Through was conceived as a program and ended up as an experiment (Egbert, 1973). As a well-funded experiment, Follow Through Programs took place in a large number and variety of urban and rural locations, encompassing diverse ethnic and cultural climates. The 160 program sites ranged from populous cities to Native American reservations. Generalization is more reasonable from any positive findings as a result of having both a large number of sites and an unusually diverse placement of these sites.

The large number of sites also permits the selection of credible sub-samples for small scale follow-up studies. (McDaniels, 1975)

The experimental nature of the Program encouraged record keeping. The activities of many of these sites are well documented in historic federal records and in sponsor records. Extensive data archives exist describing Follow Through children. However, these archives are aging and are fragile. Many of these data files are being destroyed by age, disinterest, and changing technology. If studies are to be initiated, the field should be encouraged to protect these records for a few more years.

The yearly increases in funds for the program allowed new students to be added each year. Follow Through had successive cohorts of students who entered each year. The cohort nature of the program allows preliminary studies to be conducted as well as follow-up studies to confirm findings.

Follow-up Design Strategy

The size of the Follow Through Program provides numerous design opportunities. There are so many sites that a follow-up study can be pursued as a series of special studies. When some changes in the initial study of Follow Through were made in

1971-72, the evaluation became a number of smaller studies (McDaniels, 1975). It was possible, for example, to carve out a study of large urban cities because so many sites existed in Philadelphia and New York.

There are advantages to having a follow-up comprised of several small studies. First, small studies can be understood and managed. Second, disasters can be contained. Third, things learned in one study can inform others. Fourth, resources can be spread across fiscal years. Fifth, studies performed independently of those funded directly can be incorporated in the organizing report.

In the Follow Through evaluation, redesigned in 1970-71, this design strategy was intended to improve site selection rationale, reduce data collection, and make analyses easier.

In the follow-up study, this design strategy would allow the investigators to determine what it is important to learn and then use the most reasonable evaluation strategy for individual questions. Case studies and surveys, for example, could both be used where appropriate. Mary Kennedy's work in the evaluation of the Education for All Handicapped Act (DHEW, 1979) and in the evaluation of Chapter 1 (DOE, 1986) are outstanding examples of such a strategy at work.

A sequence of studies will allow evaluation of the problems associated with finding the sample, develop a basis for estimating the costs of different data collection strategies and identify previously unanticipated threats to the eventual interpretations of any findings.

Finally, the concurrent funding of a number of small studies allows mobilization of a number of contractors with unique expertise for the work needed. Follow Through sponsors, for example, have existing data bases that would be a major resource for some study questions.

The Program's Participants are Becoming Young Adults

Ultimately, the success of the Follow Through Program was to be measured by the future "life success" of its participants. The early research design focused on proxies for the hoped-for outcomes such as early academic outcomes, self concept measures, thinking skills, and locus of control measures.

The initial Follow Through students are now 18-22 years old, and additional cohorts will be added to this "young adult" age group during the next several years.

As a result, the variables measured no longer have to be proxies for the hoped-for outcomes. We can measure the variables of life success directly.

What Outcomes Might be Found?

The age of the participants allows examination of the real goals of the Follow Through program such as frequent employment, reasonable health, and positive social values. These variables have never been studied. Any reasonable study would also include variables and measures used in other studies of the general population of young adults.

BEST COPY AVAILABLE

For example, in the report, "The Forgotten Half: Pathways to Success for America's Youth and Young Families," indicators such as the following were used:

1. real median income of families by age of family head;
2. primary families with income below the poverty line by age of family head;
3. percent of 20-24 year-old males with real annual earnings at or above the three-person poverty line by race/ethnic group; and
4. proportion of 20-24 year-old males who were married and living with their spouses by annual earnings and by educational subgroups.

Another example would be the variables in Lawrence Schweinhart's report, "The Cost-Effectiveness of Early Intervention":

1. frequency of placement in expensive remediation/special education programs;
2. frequency of high school graduation/GED;
3. frequency of post-secondary training;
4. employed/not employed;
5. self-supporting/not self-supporting;
6. arrested/not arrested; and
7. healthy/not healthy.

By beginning with the measures of others who have made significant efforts to describe the young adult population, the follow-up study of the Follow Through population will have a sound basis for comparison.

This paper focuses on the changes that were forecast for the children. Numerous accounts have been given by Follow Through participants of the changes that occurred in teachers, in parents and in the school communities where Follow Through programs were operating. Recommendations for documenting these changes will be made and are reasonable. However, the decision to proceed with a follow up study must be based, first, on the importance of studying the Follow Through children as young adults.

Should Comparison Groups be Created?

The use of the variables described above permits comparison to statistics coming from other major studies, as opposed to a constructing a peer comparison group.

Attempts to construct comparison groups for studies done while the Follow Through participants were still in elementary school produced endless problems.

BEST COPY AVAILABLE

There were always disputes regarding the meaning of differences between the comparison groups and the Follow Through groups.

Because of the discrediting confusion surrounding the use of comparison groups, the measurement instruments used in the early studies were changed. The young Follow Through students were eventually tested with nationally normed instruments and compared to national norms. This ended battles over the quality of comparison groups. Any variables measured in this study should be commonly examined attributes found in other studies of similar populations.

Comparing Individual Sponsors Should be Avoided

The Follow Through Program's sponsors were distinguished by hard work, tireless support of the school sites in which they were working, good ideas and good intentions. The authors of these ideas and intentions would obviously vote to create an evaluation design that follows-up on their work.

The leadership of these programs spoke frequently about the difference of their approach as compared to the others. There will be appeals to have the impacts of the different sponsors be compared, once again, in these studies. This should be avoided.

The important feature of the sites lead by all sponsors is that talented and committed leadership was brought a considerable distance to provide coherent, carefully implemented education programs. Each of these groups were committed to Follow Through's principles of child worth, parent empowerment, teacher involvement and training, carefully planning, and sequential instruction in the K-3 program. However, for the variables cited for study in this paper, the sponsors were not the Follow Through program. Sponsors were a component in a complex service program.

Summary

The Follow Through Program is an ideal candidate for a follow-up study. Its purpose was to make a difference in the lives of the participants, not only as students, but also as adults. It was a rich program in terms of its mission, its leadership, and its financing. It had social impacts documented at the time. The questions is how frequently impacts occurred and how lasting the impacts might be.

Finding that the children who experienced the Follow Through Program are more successful in their young adult years would be significant and find wide acceptance. Such a finding does not require a revolution in our thinking. Such a finding supports our hopes for schooling.

Finally, positive findings can lead to a realistic and powerful renewal program. The primary components of the Follow Through Program are well known and accepted as reasonable components of empowerment programs. Positive findings would energize and direct the already existing capacities of local communities and higher education.

References

Barnes, E. J. (August, 1973). *Parent Participation: An Integral Facet of Successful Educational Strategies.* Paper presented at the American Psychological Association Annual Meeting.

Becker, W. C., & Englemann, S. (1978). *Analysis of Achievement Data on Six Cohorts of Low Income Children from 20 School Districts in the University of Oregon Direct Follow Through Model.* Unpublished manuscript, Follow Through Project, Technical Report #78-1. Eugene: University of Oregon.

Cline, M. G. et al. (March, 1974). *Education as Experimentation: Evaluation of the Follow Through Planned Variation Model; Volume IA: Early Effects of Follow Through.* Final Report, OEC-0-72-5221; Office of Planning, Budgeting, and Evaluation; U.S. Department of Education.

Egbert, R. L. (April, 1973). *Planned Variation of Follow Through.* Paper presented at the Brookings Institution Panel on Social Experimentation.

Haney, W. (1977). *Reanalysis of Follow Through Parent and Teacher Data.* Boston: Huron Institute.

McDaniels, G. L. (1975). Evaluation Problems in Follow Through. In A. Rivlin and M. Timpane (Eds.) *Planned Variation in Education.* Washington, D.C.: Brookings Institution.

Rhine, W. R. (August, 1973). *Follow Through: A Model for the Utilization of Behavioral and Social Sciences in Solving Social Problems.* Paper presented at the American Psychological Association Annual Meeting.

Schweinhart, L. J. (April, 1983). *The Cost-effectiveness of Early Intervenion.* Paper presented at the Parents Role Conference, Washington: National Institute for Handicapped Research.

U.S. Department of Education (1986). *Poverty, Achievement and the Distribution of Compensatuory Education Services.* Washington, DC: U.S. Department of Education.

U.S. Department of Health, Education, and Welfare: Office of Education (1979). *Progress Toward a Free Appropriate Public Education: A Report to Congress on the Implementation of Public Law 94-142: The Education for All Handicapped Children Act.* Washington, DC: U.S. Department of Education.

BEST COPY AVAILABLE

Lessons Learned

Russell Gersten
Wesley Becker
University of Oregon

The Role of External Change Agents in Effecting School Reform

Follow Through involved extensive and intensive collaboration between poor schools, federal project officers, and sponsors in universities and educational laboratories within the context of a set of federal regulations called the *Follow Through Guidelines*. The introduction of program model sponsors, project officers, and the Follow Through Guidelines as change agents were unique events at that time in the attempts by government to improve schools in poverty areas. Yet, because of limitations in the research designs utilized in the National Evaluation of Follow Through, little systematic inquiry was conducted on the "process of change" that occurred because of these special relationships.

The purpose of this essay is to discuss potential issues that could still be explored. The framework for this discussion is provided by a recent essay by Milbrey McLaughlin (1990), which synthesizes current research on the change process. Her essay is a useful starting point because it not only discusses findings on the change process, but also includes issues that remain unresolved or perplexing. Thus, it serves as an excellent point of departure for a discussion of contemporary research documenting the processes of school change or reform.

This essay will present an overview of current thinking about the process of change using McLaughlin's framework. The points she raises will be elucidated by examples from our own research (Gersten, Carnine & Williams, 1982; Gersten, Carnine, Zoref & Cronin, 1986) on the implementation process of the Direct Instruction Follow Through Model, research on implementation of direct instruction conducted by other researchers (Cronin, 1983; Emrick & Peterson, 1979; Stallings, 1975), as well as our own experiences. The essay will also delineate promising areas for policy research, and provide a context for such research.

In the next section, we will review McLaughlin's major findings and insights and discuss their relevance for Follow Through research and for national policy. Her material has been reorganized for several reasons. McLaughlin's essay is a retrospective look at the findings of the original Rand Change Study (Rand, 1975, 1977), which was extremely pessimistic about the chances of outside change agents successfully effecting change in schools. She reexamines some of the findings from that study on the basis of more recent research such as the DESSI study (Crandall, et al., 1982), the policy research of Huberman and Miles (1984), the synthesis of Michael Fullan (1982), *The Meaning of Educational Change*, and her own more recent research on Chapter 1 (Elmore & McLaughlin, 1988; McLaughlin et al, 1985). All of this research led her to conclude that some of the earlier, pessimistic findings of the 1970s were erroneous. In her essay, she delineates 1977 findings that still hold true, as well as those that need to be rethought, revised, or reconsidered.

Major Findings from Research on Innovation

The original finding from the Rand Report, that many innovative ideas and models are adopted in name only, still remains relevant. McLaughlin notes that "adoption...did not insure successful implementation. Project resources (money, staff) did not predict outcomes....Resources alone did not secure successful implementation or project acceptance" (p.12).

She concluded, "What a project was mattered less than how it was carried out." This finding reiterates one of the major conclusions of the independent evaluation of Follow Through conducted by Abt Associates (Stebbins, et al., 1977). The Follow Through evaluation found great discrepancies in how successful a given model was from community to community.

Because of the multi-site nature of implementation of all the major Follow Through models, studies relating implementation strategies to outcomes could be valuable.

McLaughlin identifies the major factors that the Rand Report concluded were the ingredients of successful implementation. Two of these factors include:

1. extended training that is concrete and teacher-specific, and
2. classroom assistance, not only from the outside change agent, but also from some local program facilitator. (See also Hord, Rutherford, Huling-Austin, & Hall; 1987.)

Follow Through was predicated on the belief that serious change requires ongoing professional development activities that dealt with the individual needs of each classroom teacher. Yet the national evaluation did not seriously look at

BEST COPY AVAILABLE

implementation (i.e., it did not document the manner by which the various projects conducted professional development and classroom assistance activities). Serious inquiry into this matter is essential, because a good deal can be learned about strategies that are successful and those that are problematic.

This line of inquiry can build on the research already conducted on implementation of the Direct Instruction Model (Gersten, Carnine, Zoref & Cronin, 1986; Gersten, Darch, Davis & George, 1990; Gersten, Carnine & Williams, 1982) and the ALEM Model (Wang, et al., 1985; Leinhardt, 1977).

In particular, we could examine the differences between approaches that provide concrete and specific feedback to classroom teachers, versus those that see the role of the program sponsor as facilitative and collaborative. Although McLaughlin found that those approaches that are concrete and specific are more successful in implementation, many would argue that reflective, facilitative, collaborative approaches are superior (see for example, Shulman, 1987 & Schon, 1983.)

Careful inquiry about Follow Through implementation could flesh out the constructs in McLaughlin's findings. By conducting observations of current ongoing Follow Through models in operation, performing retrospective interviews, and using supporting file data with those involved in professional development activities, we could begin to better understand which factors lead to successful professional development in a detailed way. The work of Bruce Joyce (1990) on coaching could serve as a framework for examining the nature of the sustained interactions between classroom teachers and the representatives of the models.

McLaughlin links success to projects that succeed in providing "the on-going and sometimes unpredictable support teachers needed" *(p. 12)*. Serious inquiry can be conducted on how certain universities or laboratory sponsors were able to provide this type of support.

Research can also spell out the role of local program facilitators, explaining how these individuals were trained, and whether they utilized their experience as program facilitators to move on to other leadership experiences in the district.

Another key finding by McLaughlin is that even long-term implementation did not predict institutionalization. Follow Through is very fertile ground for this key policy issue. We could explore the extent to which districts institutionalized certain Follow Through procedures and strategies, and the reasons why.

For example, our own research in Flint, Michigan demonstrated that even when the Follow Through program ended and the highly specialized Distar curricula were replaced with traditional basal reading and mathematics series, classroom teachers transferred to their new materials many of the effective teaching techniques and strategies they had acquired (Kinder, Gersten, & Kelly, 1988).

In 1980, the San Diego Unified School district was searching for a system to improve the quality of educational services in 28 schools with predominantly minority, low-income students. Because of the documented success of the Direct Instruction Follow Through program in seven schools, the district decided to borrow some components of the Direct Instruction Follow Through Model. In particular, they implemented the system for curriculum-based assessment of student progress, the monitoring of curriculum pacing, and the use of directed lesson guides that highlighted key lesson objectives and provided systematic practice and review on

BEST COPY AVAILABLE

targeted skills and strategies. In accordance with the Follow Through model, they hired and trained a school-level program facilitator for each school to assist with curriculum implementation. This exemplifies what Berman and McLaughlin called *mutual adaptation.*

We are reasonably sure there are many other instances of institutionalization and mutual adaptation of ideas, practices and policies originated as part of local Follow Through programs, although no serious inquiry has been conducted to date. This would be an excellent research area for those interested in understanding policy in terms of procedures for introducing innovative educational practices into the schools, and for understanding the subtle interplay between the agenda of districts and the ideas and practices presented by universities and educational laboratories.

External school reform programs such as Follow Through do introduce new sets of ideas, systems for providing professional development, and teaching procedures. Examining this legacy in depth would provide rich information as to how ideas diffuse through school districts and how they are transformed.

There is another side to the legacy of Follow Through that has rarely been discussed in a formal fashion but needs to be told—that is, the role of Follow Through in facilitating state, federal, and local policy on racial integration, and in creating situations where parents could play an active role in ensuring that their children receive high-quality education.

Many of the sponsors played an active role in this process, especially in the early years (the late 1960's and early 1970's). We are certain that the lessons learned during that era were valuable, and need to be retold, analyzed, and synthesized. A recent essay by Aragon (1989) discusses integration of East Las Vegas, New Mexico schools as a result of the Follow Through experience. When Gersten visited Uvalde, Texas schools in 1986, he noted that half the teachers in one school had been parents of Follow Through students, and that many had begun their careers as educators by serving as paraprofessional aides in Follow Through, going on to receive teaching credentials as a result of this training. The degree of commitment to the instructional program and the awareness of the educational needs and community situations was astonishing. Uvalde schools were mostly white directed and taught when we first came there in 1968. Similar events occurred in Racine, Wisconsin; Mission, South Dakota; Flint, Michigan; and East St. Louis, Illinois. Inquiry into the role of Follow Through sponsors, and of external change agents in general, as institutions which promote racial integration, and the empowerment of parents and other minority group members through their active role in the educational process, is essential.

One of McLaughlin's major findings of great relevance to current policy debates deals with project scope. The Rand study of the mid-1970's concluded that projects that tried to tackle everything simultaneously by attempting to create system wide change all at once rarely succeeded. She concluded then, and sees nothing in subsequent research that would lead her to change her conclusions 15 years later, that:

> *Planned change efforts, it seemed, needed to be sufficient in scope to challenge teachers and kindle interest, but not so ambitious that they required too much too soon from the implementing system (p. 12).*

BEST COPY AVAILABLE

In this era of endless calls for system-wide overhauls of curriculum content, pedagogical procedures, grouping, management of schools, and the nature of community-school relationships, it is essential that this finding be reexamined and reexplored.

The variety of Follow Through models is an excellent venue for such study. The major sponsors approached change from very different vantage points. Some, such as the University of Oregon and the University of Kansas, attempted to slowly, but steadily, improve the technical aspects of classroom instruction. Their involvement with districts focused on how the curriculum was taught, with extensive attention to the training of teachers in order to more carefully assess student progress through the curriculum, to present adequate, clear examples of new concepts, to sufficiently motivate students, etc.

Other Follow Through models attempted change of a much broader scope. These included efforts to radically alter the content of the curriculum, the structure of classrooms, and the way teachers think about teaching and children. Such approaches included the High Scope model which stressed the conceptions of Piaget, and the Bank Street Model which stressed the insights of psychodynamic thinking and child development theory. Still other models attempted to radically alter the relationship between homes and schools, such as the Far West Laboratory, the Open Education Model, University of Arizona, and the Southwest Lab.

Comparisons of the relative successes and failures of these approaches, and attempts to link these successes and failures to the scope of the projects, could have profound implications for the wave of university-school district collaborative efforts that have begun to burgeon in the 1990s.

The next finding, too, has profound policy implications: "the enthusiasm engendered in teachers may come to little because of insufficient will or support in the broader organizational environment, which is hard to orchestrate by means of federal (or even state) policy...although teachers in a site may be eager to embrace a change effort, they may elect not to do so, or to participate on only a pro forma basis, because their institutional setting is not supportive" (p. 213).

Follow Through provides an excellent venue for an examination of why certain practices lasted and flourished in certain communities, and why they languished, or were implemented only in a "pro forma" basis, in others. Case studies could help us understand the factors that lead to building credible, meaningful, active support at the local school district level. By pooling together case studies of relative successes and failures, we are likely to develop some reasonable hypotheses and insights into the nature of effective district-university/laboratory relationships that could help those involved in this endeavor in the 1990s.

Some pragmatic issues could be pursued in terms of delineating the details of how active commitment at the school district level is developed and maintained. This can include descriptions of the day-to-day-work of the Follow Through local program facilitators, such as was done in the exploratory research by Gersten, Darch, Davis, and George (1990). This also could include interviews with these individuals and the teachers they served, to better understand which functions were perceived as most beneficial.

BEST COPY AVAILABLE

Material such as that generated by Wang, et al., (1985), and Gersten, et al., (1986), which displays examples of the forms local program facilitators used to assess how well the Follow Through model was being implemented, could serve as a starting point.

McLaughlin concludes with two findings from the 1975 Change Agent Study that no longer hold up. She provides some methodological reasons for the earlier, somewhat erroneous findings, and discusses more recent research that supports her current position on the two issues.

Both of these "revised findings" have important policy implications, and, again, serious inquiry into Follow Through can help shed light on misconceptions which have remained in the field of educational policy for many years.

McLaughlin first asserts that she and her colleagues were wrong in assuming that a teacher's initial motivation to participate in a school reform project was an important predictor of how successful the project was. After her subsequent research on Chapter 1 implementation (the research of Guskey and Crandall), she now concludes that, quite often, "...belief follows practice. Individuals required to change routines or take up new practices can become believers" (p.13). This issue is also explored in depth by Guskey (1986), and Gersten and Guskey (1985).

Current research indicates that a whole array of factors—the clarity of the innovation, the level and quality of professional development, and the extent to which teachers see visible impact on their students (especially their difficult-to-teach-students)—are far more important than initial teacher motivation or attitude, Gersten & Miller, 1987.

This would be an excellent area for research involving Follow Through. Typically, Follow Through was introduced into a district by some combination of parents and administrators. Rarely did teachers have much initial say in the selection of a Follow Through approach. Yet there are many, many instances of teachers who became staunch "believers." Understanding the process of how teachers' beliefs evolve and helping to clarify some past misconceptions would have profound policy implications.

McLaughlin's final misconception of 1975 was that, overall, outside change agents such as universities could have a serious impact on changing the nature of schooling. She now feels that she and her colleagues overstated the case. The truth is that universities and outside change agents rarely have a long-lasting impact on school districts because they rarely provide the type of consistent professional development activities and troubleshooting required for lasting change.

Because of its nature, Follow Through can well serve as the exception that proves the rule. There are many instances of successful long-lasting involvements between districts and universities. Helping to articulate the reasons for this could and should be a major area for policy research. The eight issues listed above will all flesh out this issue.

Methodological Issues in Conducting Longitudinal Research

The purpose of this paper is to discuss methodological issues and procedures for conducting longitudinal research on the subsequent life experiences of students

who participated in the Follow Through programs in the 1970's. It is based, in large part, on our experience conducting four such studies.

This paper will deal with issues involving the selection of comparison groups, data collection, measures, and attrition. We will also discuss some of the intricacies involved in establishing and maintaining the district cooperation and support of this type of research. We begin by addressing an issue which plagued the earlier evaluation of Follow Through conducted by SRI and Abt Associates: the selection of valid comparison groups.

Selection of comparison groups

The original Follow Through evaluation has been criticized because in many communities, the Follow Through sample and the comparison group were not equivalent. In several cases, the comparison schools were those left over after the poorest schools in the district were placed in Follow Through. Certainly great care needs to be taken in selecting comparison schools for this type of longitudinal research. Even in instances where SRI found a reasonable comparison school for a Follow Through program, we were unable to link up to the to the SRI data base, since some has been destroyed or all the ID's removed.

In contacting school districts to locate demographically similar comparison schools, researchers need to be aware that district personnel are likely to provide them with a school that is demographically similar in 1991. However, one must ascertain that these schools were similar in 1970-74. Some archaeological work may be required. We found that, in many instances, school populations had shifted dramatically over the years. In some cases, this was due to demographic trends; in other cases, this was caused by activities related to integration—busing and/or magnet schools.

The AFDC (Aid to Families With Dependent Children) data for 1970-1972 seem the most reliable for finding demographically similar schools. We were able to access these data in all six communities. We tended to utilize these data, coupled with data on ethnicity, to find the most suitable comparison schools. In some cases, we were able to access achievement test score profiles of the schools in the years prior to Follow Through. This served as an additional indicator of comparability.

However, in some communities, no school was comparable to the Follow Through school. In these cases, we thought it better to use an interrupted time series design than to compare Follow Through students to those from a different culture or income level. For example, in the Cherokee, North Carolina, community, Follow Through was implemented in the Bureau of Indian Affairs school, where all students were Native American. The nearest school was 20 miles away, and the population was primarily Anglo. Another example is Uvalde, Texas, where the Follow Through children were exclusively Hispanic. The only possible comparison group was totally Caucasian. In both instances, it seemed advisable to merely describe trends over time, rather than make formal comparisons.

Criteria for Including Children in a Comparison Group. In our studies, we decided to include all children who had at least three years of Follow Through, and a child in the comparison schools who spent at least three of the years (during K-3 or 1-3) in that comparison school.

Let's say that Murphy Elementary was the comparison school. We included all the children who had at least three years in the Follow Through program during 1-3 or K-3 in Murphy Elementary. We would have preferred to require four years (K-3), but it lowered the sample size dramatically, and we reasoned that in three years, children in Follow Through should have had a lengthy enough exposure to benefit from the program. At least three full years in either the Follow Through program or at the comparison school during the K-3 years seems a reasonable criteria.

The nature of the instructional program in the comparison school needs to be seriously investigated. We learned the importance of this aspect in Flint, Michigan. As the research staff was reviewing the cumulative records of students in the comparison schools, they noticed DISTAR next to many students' names. A bit of detective work revealed that, due to a grant from the Mott Foundation, about eight schools in Flint were using the DISTAR direct instruction program in the primary grades for a two-year period. The years of implementation were 1970-72, exactly the years of the proposed longitudinal evaluation. We needed to find another comparison school which was not implementing direct instruction.

Data collection procedures

Our experiences in the six communities in which the research was conducted have proven that this type of longitudinal research can be done. Students can be tracked from elementary school through high school, and relevant data collected. However, this research can be exceedingly difficult, painstaking, and, at times, frustrating.

For example, the study in New York City that we conducted with Dr. Linda Meyer traced students from one Follow Through school and one local comparison school over a 13-year period. Over the course of the years, students from these two elementary schools in Brooklyn had dispersed to approximately eighty different high schools in all five boroughs. Many had transferred from one high school to another. Contact with all eighty high schools was necessary to complete the data set.

Even though many districts now utilize sophisticated computerized systems for the children currently in their systems, few such systems were in use in 1970 and 1971. In every case, we found that we needed to manually trace the students.

The percentages of students retained a grade or more in the schools we investigated was large. District files are set up according to graduating classes, not according to when students began kindergarten or first grade. In order to trace the group of students who began kindergarten in 1970, for example, one would need to cast a broad net over graduating classes—not only looking at the graduating class of 1983, but at those of 1984, 1985, and 1986. In addition, the researchers must carefully review the "drop" or dropout files for each of these graduating classes.

The collection of precise information on dropouts is a vast problem. Not only does each community define the term differently, but records often are not systematic. Terminology is complex and confusing, and varies from state to state. In New York City, for example, the word dropout is never used—the code word is "17+". In some states (Oregon, for example), children's cumulative records are placed in the "drop" file if they leave the school for any reason, including excessive absences, being dropped from the rosters (the common definition), officially telling the counselor they no longer wish to attend school, or moving to a school in another county.

BEST COPY AVAILABLE

Rarely does an adolescent come into school, shake the principal's hand, and say, "I'd like to drop out." They more often simply stop coming to school, and their files are sent to a truant officer or counselor. Some files remain in a limbo state. This means that each case must be thoroughly investigated. Researchers must review the data of all students who moved to see whether the student actually attended the designated school.

Measures

Rather than providing suggestions for a set of measures, which was done in the position paper by the Denver planning group, this section will merely discuss a set of pragmatic issues related to the measures collected in the longitudinal research conducted on the Direct Instruction students. In this research, we attempted to collect information on achievement test scores at grades 9 and 11 (years when many districts conduct standardized testing), grade point average, whether the student was retained during any of the years after Follow Through, special education placement, whether the student dropped out of school or graduated, attainment of GED (high school equivalency) by dropouts, and acceptance into college.

Ninth grade achievement is a viable measure to examine, because most districts test students at that level. However, collecting these data requires the "broad sweep" across four years of testing described earlier to obtain scores for all students (even those who were retained). In other words, if the student began kindergarten in 1970 and therefore was "scheduled" to be in the 9th grade in the spring of 1981, one must also review the 9th grade testing results for 1982, 1983, and even 1984. With this procedure, we were able to locate achievement test scores for most students, even those who were retained. Typically obtaining ninth-grade test scores from this era was not a difficult task, because districts tend to keep records of standardized testing for a long time.

We were much less successful in obtaining standardized test scores for grades 10, 11, and 12. There are several reasons for this problem. We found that standardized achievement tests are not given in these grades in some districts. The fit of the high school curriculum to the content of standardized achievement tests is much weaker in high school. Even in districts that gave such tests, we noted that quite a few of the students in the Follow Through and comparison samples never made it through 11th or 12th grade, due to dropping out of school. Finally, we noticed that quite a few of the very low-achieving children who were registered in school did not show up on 11th grade achievement test rosters. Since this was a recurrent phenomenon, we began to wonder whether these students tended to cut school on test days.

There were so many missing data points at this level that we felt it was best to exclude these data. For all these reasons, 10th and 11th grade achievement data will be difficult to obtain, and may provide little valuable information.

Graduation/dropout rates. These data are always available. As previously mentioned, a "sweep" of up to four graduating classes may be required, and cooperation between several different offices within the district may be necessary. The information is always somewhere in the schools, but this is very intricate, sensitive data to collect. Researchers must realize that definitions will vary from place to place, and that it's a fairly taboo topic.

It is useful to measure the percentage of students who graduated "on time" versus those who graduated one or two years later. In one of our studies, there was a significant effect in one of the communities. Also note that in some districts, high school students are not officially retained; they simply don't graduate until they have earned enough units.

High school equivalency. Determining whether or not students at this stage completed their GEDs will, in all likelihood, require direct contact with the client, since school districts do not seem to have this information. Typically GEDs are completed through either night school or a community college. Therefore, this measure requires a good deal of linkage with other community agencies.

Retention measures are, again, possible to collect. This entails the manual tracking of each student's cumulative record. One needs to be aware that in many communities, retention policies vary dramatically from decade to decade. In the early and middle 1980's, there was a large scale movement to retain children who had not mastered academic content. Currently, there is a widespread movement to retain as few students as possible. Therefore, one must remember that this is a variable which heavily reflects policy shifts. Many Follow Through models, including Direct Instruction, did not believe in this retention, and this may have an impact on retention rates.

Special education placement was a key variable in Larry Schweinhart and David Weikart's research on the later effects of the High Scope preschool program. We found little data on this topic. Essentially, almost none of the Follow Through or comparison students were placed in special education in any of the six communities.

We concluded that we could not therefore assert that virtually none of these students needed special assistance, or that virtually none would be referred into special education using policies and procedures of the 1990s.

The era we were investigating was prior to implementation of Public Law 94-142, requiring free and appropriate special education services to all students in need. In the low-income schools we were investigating, it is likely that free special education services were rare prior to 1978. Thus, while the students were in the elementary and middle school grades, only students with severe problems were likely to be referred. In addition, it was a policy in the Direct Instruction model, and probably some of the other models, to rarely use special education placement.

Attrition

In each of the former Follow Through sites, even those with extremely high mobility, we were able to obtain a sample of at least 20 or 25 students who remained in the district over the entire 13 years of the investigation. The sample sizes were always adequate for some types of statistical analysis.

In one study, we determined if there was any bias in the sample of students who remained in the school district. (Keating & Gersten, 1987). It compares the third grade achievement of those students who are stayed in the community until high school with third grade performance of those who moved out of the community.

BEST COPY AVAILABLE

We found that during grades 4-12, in five of the six cases these groups were roughly comparable. In only one of the six cases, the children who stayed in the community were somewhat higher in third grade achievement than those who left.

District linkage and support

This essay concludes with a discussion of an issue that is more psychological and political than technical. Many of the students we plan to research attend some of the poorest, lowest-achieving schools in the city. When discussing these low schools with school district personnel (indicating that the research is to examine not only achievement and college admissions data, but also drop-outs and retention), one is likely to find some defensiveness among some school personnel, even if it's unconscious. At times, individuals may provide evasive information, even misinformation. This defensiveness is in part due to the fact that we are examining programs and policies which existed long ago, as well as chronic problems which have experienced numerous policy and procedural shifts over the years. In addition, we need to confront the fact that what we call a "comparison" school represented the typical educational program that the given community provided—a program that in some cases was based on extensive input from local district personnel.

In Follow Through, each of the sponsors worked primarily with elementary education personnel (K-3) and compensatory education personnel. The longitudinal research proposed requires involvement with secondary education personnel (who typically do not even know that a program such as Follow Through ever existed in the primary grades) as well as coordination with counseling, attendance and guidance personnel, and the district research and testing personnel.

The proposed research requires an interface with many facets of district operation, and the liaison is massive. After conducting six such studies, we concluded such interface is usually possible to maintain. We were successfull in conducting these studies in both very small, rural communities, such as Williamsburg County, South Carolina, and in very large, urban centers, such as New York City. Executing and completing these types of longitudinal studies is possible, but it is much more complex than one might think, requiring the tactful building of relationships and perhaps reinforcers for those whose time will be involved.

References

Aragon, M. (1989). An historical perspective of the early days of the follow through in Las Vegas, New Mexico. *Direct Instruction News*, 9(1), 5-6.

Crandall, D. P., et al. (1982). *People, policies and practices: Examining the chain of school improvement* (Vols. 1-10). Andover, MA: The NETWORK Inc.

Cronin, D. P. (1983). *Personal and organizational factors affecting educational program implementation: A case study*. Ann Arbor, MI: University Microfilms International.

Elmore, R. F., & McLaughlin, M. W. (1988). *Steady work: Policy, practice, and the reform of American education*. Santa Monica, CA: Rand Corporation.

Emrick, J. A., & Peterson, S. M. (1979). *Implementation study: Case study report and first-year measurement development: Pt. I. Introduction and case study report.* Los Altos, CA: Emrick.

Fullan, M. (1982). *The meaning of educational change.* New York: Teachers College Press, Columbia University.

Gersten, R., & Keating, T. (1987). Improving high school performance of "at risk" students: A study of long-term benefits of direct instruction. *Educational Leadership*, 44(6), 28-31.

Gersten, R., & Miller, B. (1987). The study of implementation: Implication of Follow Through Research. In M. C. Wang & E. A. Ramp (Eds.), *The National Follow Through Program: Design, implementation and effects.* (Final Project Report, Vol. 1). Philadelphia, PA: Temple University.

Gersten, R., Carnine, D., & Williams, P. (1982). Measuring implementation of a structured educational model in an urban setting: An observational approach. *Educational Evaluation and Policy Analysis*, 4(1), 67-79.

Gersten, R., Carnine, D., Zoref, L., & Cronin, D. (1986). A multifaceted study of change in seven inner-city schools. *Elementary School Journal*, 86(3), 257-276.

Gersten, R., Darch, C., Davis, G. & George, N. (1991). Apprenticeship and intensive training of consulting teachers: A naturalistic study. *Exceptional Children*, 57(3), 226-236.

Gersten, R., Guskey, T., & Cronin, D. (1985). *Transforming teacher reluctance to teacher commitment.* Paper presented at the annual conference of American Educational Research Association, Chicago, IL.

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher*, 15(5), 5-12.

Hord, S. M., Rutherford, W. L., Huling-Austin, L. & Hall, G. (1987). *Taking charge of change.* Alexandria, VA: ASCD, Southwest Educational Development Laboratory.

Huberman, A. M., & Miles, M. B. (1984). *Innovation up close: How school improvement works.* New York: Plenum Press.

Joyce, B. (Ed., 1990). *Changing school culture through staff development.* Alexandria, VA: ASCD.

Kinder, B., Gersten, R., & Kelly, B. (1988). *Expert mathematics instruction for at risk minority students* (March, 1989). Paper presented at annual conference of American Educational Research Association, San Francisco, CA.

Leinhardt, G. (1977). Program evaluation: An empirical study of individualized instruction. *American Educational Research Journal*, 14(3), 277-293.

McLaughlin, M. (1990). The Rand change agent study revisited: Macro perspectives and micro realities. *Educational Researcher*, 19(9), 11-16.

McLaughlin, M.W., Shields, P.M., & Rezabek, D.J. (1985). *State and local response to Chapter 1 of the Eucation Consolidation and Improvement Act, 1981* (Technical Report No. 85-A6. Stanford: Institute for Research on Educational Finance and Governance.

Rand Corporation (1975). *Volume V: Executive summary* (Research Report No. R-1589/5-HEW). Santa Monica, CA: Author.

BEST COPY AVAILABLE

Rand Corporation (1977). *Volume VI: Implementing and sustaining Title VII bilingual projects* (Research Report No. R-1589/6-HEW). Santa Monica, CA: Author.

Schon, D.A. (1983). *The reflective practitioner.* New York: Basic Books.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review,* 57(1), 1-22.

Stallings, J. (1975). *Follow Through program classroom observation evaluation.* Menlo Park, CA: Stanford Research Institute.

Stebbins, L., St. Pierre, R. G., Proper, E. L., Anderson, R. B., & Cerva, T. R. (1977). *Education as experimentation: A planned variation model* (Vols. IVA-D). Cambridge, MA: Abt Associates.

Wang, M.C., Genarri, P., & Waxman, H.C. (1985). The adaptive learning environments model: Design, implementation, and effects. In M.C. Wang & H.J. Walberg (Eds.), *Adapting instruction to individual differences* (pp. 191-235). Berkeley, CA: McCutchan.

Implications for Teaching

Mary M. Kennedy
Michigan State University

As I understand our roles in this conference, we are all to consider what a follow-up study of Follow Through graduates might look like, and we are all to consider some other particular issue as well. My particular task, in addition to the general task, is to consider the implications of findings from such a study for teacher education. To address either of these issues—that is, what such a study might look like and what its implications might be for teacher education—I need first to revisit the models that are so central to Follow Through. The notion of models is central to Follow Through and will continue to be central in any further research that might be conducted on Follow Through. The models are important to any decisions about what a follow-up study might look like because they constitute different theories of how to alter childrens' ultimate life chances. And they are important to any considerations about teacher education because they all entaile major efforts to alter teaching practice.

In the next section of this paper, I will discuss the aspects of Follow Through models that are most salient to both of my tasks. Then, in the sections that follow, I will discuss, respectively, the implications of these models for teacher learning, for teacher education, for identifying evidence of life chances, and for drawing causal inferences about how they may influence life chances.

57
BEST COPY AVAILABLE

Follow Through Models

Through a variety of different funding strategies, the U. S. Office of Education and, later, the U. S. Department of Education, has encouraged the development, demonstration, and dissemination of educational models. In fact, ever since the federal government took an active interest in education, educational models have been a part of our vocabulary. But the concept of an educational model has continued to be illusory. To illustrate my point, let me offer some different definitions of the term "educational model."

1. One way to construe educational models is as *curricula*. Most textbook publishers consider themselves to be developing curricula. They define the content that should be taught and the sequence in which that content should be taught. At bottom, curricula are ways of organizing bodies of knowledge into sequences, so that children can learn that knowledge one step at a time. Many curricula are designed to introduce content in a way that will facilitates children's movement through the sequence. If Follow Through models were curricula, they would be tacitly arguing that the way we alter life chances is by providing the right content, and the right sequence of content, during the first four years of school, thereby improving the chances of learning other content later on. In fact, very few Follow Through models even had curricula when they began, though some developed curricula over time.

2. Another way to construe models is as *pedagogues*. Cooperative learning, the writing process, and cross-age tutoring are examples of models-as-pedagogues. The rationale for these models differs from one to the next. For instance, the process approach to writing is advocated as a way to foster in children an understanding of the nature of writing, of how one writes, and of what is important to know about writing. In contrast, cooperative learning is advocated as a method for promoting higher-order academic goals in a variety of different content areas, for raising self esteem, and for helping students learn to work with others (Slavin, 1988).

3. Yet a third way to think of models is as *programs*. Many of the educational models that are validated by the Joint Dissemination and Review Panel and disseminated through the national diffusion network could be called programs. They define a number of components—counseling, remedial education, after-school care, health examinations, perhaps also curricula or pedagogues, as components that are assumed to work in tandem to improve either student learning or the quality of student life. In fact, the Follow Through program as a whole represents a model of how to improve childrens' life chances that included the provision of a wide range of services, integrated and delivered through a particular agency, the school. The constellation of services, and how they were organized and delivered to children and their families, constitutes a particular model of early childhood intervention. Visions also differ, in ways that are consistent with their emphasis on what ought to be, in what they take as the goals of education. Of all the differences one can identify among Follow Through models, differences in the outcomes they sought are the most salient.

BEST COPY AVAILABLE

The Follow Through models differ in the relative emphasis they place on curriculum, pedagogy, and program, and in how prescriptive they are about these matters. They are similar, though, in that all entail some form of theory about how these components contribute to life chances, and all entail some form of vision about what schooling ought to be like. To illustrate this point, I'd like to define three broad classes of Follow Through models and show how they differ regarding both their theories of how early education could improve childrens' life chances and their visions of an ideal early childhood program.

One group of models seeks to improve childrens' life chances by increasing the specific knowledge and skills children acquired. The most important causal argument these models make is that children will do better in life if they possess a corpus of basic knowledge and skills that will serve as a foundation on which all future learning can rest: Students need to know rudimentary arithmetic, spelling, grammar, and so forth before they can learn most other things. Their future lives will be enhanced through the acquisition of these skills because their future learning will be enhanced through the acquisition of these skills. Models in this category tend to define more specific classroom practices than other models, relying on predefined curricula (eg. programmed instruction) and/or predefined pedagogues (eg. behavior modification). Notice that, tacit in this vision, are several assumptions about what contributes to one's chances in life. One assumption is that knowledge and skills are more important than, say, character or drive. Another assumption is that what we tend to call "basic skills" must be learned before one can learn more complex knowledge or skills.

Consistent with their theories, these models envision an ideal early childhood classroom as one that assures that students learn as much as possible. They place a high moral value on instructional efficiency, for they feel it is imperative that children acquire as much school knowledge as possible.

Another group of models argue that future life chances will be influenced more by one's problem solving ability or one's self-management or self-confidence than by the possession of any particular knowledge or skills. They therefore want to give children a greater sense of efficacy in school contexts, to increase their ability to reason, or to help them learn take control of their own time. These models clearly differ from those in the first group in what they assume will contribute to childrens' life chances. Follow Through models that fall into this category tend to assume that if they could alter the learning strategies of children, they would produce the kind of life-long improvements the Follow Through program called for. This is not to say that they do not believe that specific knowledge and skills are important, but rather that this knowledge will follow from, rather than precede, these other attributes. These models are far less specific than those in the first group regarding the curricula and pedagogues teachers should draw on. Instead, they have general ideas about what teachers should be striving for.

Consistent with their theory, they envision the ideal early childhood classroom as one that fosters independence in children, that encourages children to explore subject matter knowledge on their own, and where children are free to make their own decisions. Believing that the quality of their learning experiences is critical to their future self-esteem and self-management ability, these models feel a moral imperative to provide such an environment for children.

BEST COPY AVAILABLE

Yet a third group of models assume that the best route to altering children's life chances is through the family and community rather than through the school. Believing that families and community will have more impact on children in the long run, these models seek to alter families and communities in ways that make them more supportive of the children. Follow Through models in this category seek to provide parents with further education, to help parents interact with or tutor their own children, to help them interact more productively with schools, or to help them gain control over their own lives: their jobs, their neighborhoods, or their homes.

These categories of models represent three different ideas about how to alter a child's life chances. One idea is to alter what children know, thereby better enabling them to learn still other content. Another idea is to alter children's methods of interacting with schools and with school subject matter, thereby enabling them to learn other content. And a third idea is to alter the home environment in which children live, thereby giving them a better environment in which to thrive in the future. Each model not only argues that there is a causal relationship between early childhood education and life chances, but also believes schools, and teachers in particular, have a moral imperative to produce the circumstances that are most likely to yield these benefits.

Implications of the Models for Teacher Learning

Since each Follow Through model includes a theory of how to improve children's life chances, and each envisions an ideal educational approach to schooling, a teacher who is to implement a Follow Through model must embrace the ideals of the program, not just learn a few techniques. This suggests immediately a problem for teacher education, for teachers already have their own ideas about what an ideal classroom looks like and about how such a classroom contributes to children's life chances. They may think, for instance, that their most important task as teachers is to be nice to children, or, conversely, to be strict and to force them to learn specific content. Their visions of how these behaviors contribute to children's growth and development are probably not as clearly articulated as those of the model sponsors, but they are important to teaching practice in that teachers, like sponsors, have strong emotional commitments to these visions.

Where do teachers' visions and theories come from? More than any other profession, the profession of teaching socializes new members from childhood on (Lortie, 1975). Teachers have spent over 3000 days as children and as young adults observing teachers (Kennedy, in press). Their experiences are tantamount to an apprenticeship of observation. Moreover, because of their dependence, as students, on their teachers, their conclusions are invested with emotion. By the time they graduate from high school, every prospective teacher, like every other adult, has formed ideas about what counts as good teaching and what counts as bad teaching. Their ideas are probably formed more by their own emotional reactions to their teachers than they are by careful analysis of their teachers' behaviors, and the very emotionality of their visions makes them more resistant to change.

The pervasiveness of teachers' experiences during their apprenticeship of observation, both across grade levels and across subject areas, coupled with the sheer volume of time spent observing, yields in teachers (and in other adults, for that matter) a deeply entrenched *and tacit* set of beliefs about what can and should happen

in schools: about the nature of intellectual work and the nature of school subjects, about the teacher's role in facilitating learning, and about the pedagogical implications of different kinds of students.

Though sociologists have recognized the importance of the apprenticeship of observation for some time, the details of teachers' ideas, and the constraints they place on teachers' ability to entertain new ideas about teaching, was one of the most important findings of recent work at Michigan State's National Center for Research on Teacher Education (NCRTE). For the past three years, we have been following both teachers and teacher candidates through a variety of preservice, induction and inservice programs. We found that, despite the diversity of approaches to teacher education that we studied, many of these programs were unable to alter substantially the ideas teachers had when they arrived.

We also found that, for teachers to adopt a different theory of how school experiences contribute to later life chances and a different vision of what an ideal school experience would be like, teachers need to alter their ideas about subject matter, about students from culturally different backgrounds, and about their own role in facilitating student learning. Beliefs in all three of these areas are relevant to Follow Through models.

Subject matter and how it is learned

Follow Through models represent different ideas about the nature of subject matter and about what it means to "know" a subject. Those in the first group I described tend to assume that subject matter consists of specific facts and skills which students must acquire. Their goal is to assure that students acquire as much of this content as possible. Those in the second tend to assume that subject matter is a way of thinking about and dealing with situations, and that students learn subject matter by working with it, manipulating it, and interacting with it. Those in the third group tend to think of subject matter as something that students must have an emotional relationship with, and that this can only occur when their families or communities value it as teachers do.

Teachers who think about subject matter differently than a model sponsor thinks about it will have difficulty not only understanding the model, but accepting it as well. They will worry that if they implement a model, they will be depriving children of the subject matter they really need.

Culturally-different students

Follow Through models were designed originally for a specific population—poor children who normally are not well served by schools. These are not children whom teachers are likely to understand. Most teachers and prospective teachers did well in school, liked school as children, and expect their students to be like they themselves were as students. Since prospective teachers are mostly white and female, and since the composition of the American student body is becoming more and more diverse, we now recognize even more than we did in the early days of Follow Through that teachers must both know about these students and possess certain attitudes toward them. Moreover, most of them come from small, homogeneous, lower middle-class communities. They attend college in nearby communities and

BEST COPY AVAILABLE

hope to return to their home town or to a neighboring town to teach. Their exposure to students who are even marginally different from themselves is often close to non-existent. Teachers from Christian communities, for instance, may not know about Jewish holidays, teachers from white communities may not understand Black dialect, and teachers from working class communities may have never encountered students whose families are on welfare. Consequently, many of them are ill-equipped to work with the diverse range of pupils now attending K-12 schools in the United States. We found that many teacher candidates had never considered that there might be learners who respond to school subjects differently than they themselves did (Floden, in press).

These findings are particularly salient for those models that emphasize interactions with families, but they are also important to any Follow Through model, in that all of these models are designed to function with culturally diverse students, and in particular with poor students whom the American schools have not traditionally served very well. The models assume that these children can benefit from schooling. Yet, in our work at Michigan State, we found that many teacher candidates were unable to grasp the pedagogical implications of teaching children different from themselves. They were unable to move beyond the two moral imperatives of teaching—the imperative to treat all students equally and the imperative to accommodate individual differences (Paine, 1988). They recited both of these values almost as if they were mantras, unaware of the contradiction and unaware of how to implement either one. Many teachers, despite their recitation of educational mantras about all children being able to learn, do not really believe that their students can learn.

Teacher role

Perhaps of most importance to the Follow Through models is the role teachers play in creating the ideal early childhood classroom. The models differ most significantly in their views about the teachers role, but all place a great deal of emphasis on teacher role. The teachers' role is important to the model in that it is the counterpart to the students' role. Through the role they adopt, teachers also teach children to adopt a particular role. The role that children play, in turn, is likely to have some bearing on their life chances. If anything stays with them once they leave their early education programs, it is likely to be the roles they learned to play as learners. Students in many schools learn to be antagonist toward teachers, to ignore them, to manipulate them, or to engage in any number of other counterproductive strategies for managing interpersonal and power relationships. These roles may stick with them throughout their lives, and could account for the faith that so many people have in early childhood interventions.

Follow Through models in my first group want teachers to take responsibility for student learning, for presenting content, for assuring that classroom activities are appropriately paced, for assuring that students are attentive and on task. Those in the second group want teachers to give students a larger responsibility for their own learning, and to facilitate learning by providing an environment that enables students to actively engage and explore material on their own. In these models, teachers follow the students more than lead them. Those in the third group want teachers to interact more with children's families and communities, and to adapt

BEST COPY AVAILABLE

their work to fit the local culture. None of these roles is easy to enact. Each requires intense involvement, which in turn can only come from a complete understanding and embracing of the model's vision.

In our work at Michigan State, we found that undergraduates who enter teaching almost universally hold a limited view of their role as teacher, thinking that teaching entails almost nothing more than telling students what they know and assessing students' recall of that knowledge (Ball, 1988; Feiman-Nemser, McDiarmid, Melnick & Parker, in press; McDiarmid, in press). This tacit view of the teacher's role limits teachers' abilities to adopt any of the Follow Through models. Having observed teachers throughout their lives who taught mainly by telling students and then testing students, teachers cannot imagine the effort that is implied by any of the Follow Through models to alter children's learning and hence their life chances.

More than any other aspect of instruction, teachers must adopt a role and make it their own; they cannot implement Follow Through models simply by following a procedural manual that outlines their activities. A teachers' ability to adopt the role appropriate to a particular Follow Through model depends not only on his or her understanding of that role, but also of her emotional acceptance of and commitment to that role.

Moreover, helping novices grasp a specific role may be the most difficult challenge facing teacher educators. Role is an abstract concept to grasp, and the roles envisioned by Follow Through models must be adopted cognitively, effectively, and kinesthetically by teachers. Without such a complete adoption, teachers will simply not be able to enact the role that Follow Through model sponsors envision.

Implications of the Models for Teacher Education

Follow Through sponsors did, of course, engage in a great deal about teacher education as they tried to help teachers learn to implement the models they espoused. And there were many reasons to expect their efforts to be successful. Unlike preservice teacher education, for instance, they were working with experienced teachers, who already had enough classroom experience to be able to grasp the meaning and significance of sponsor messages in a way that preservice teaching candidates rarely can. Second, they worked with teachers in their own classrooms rather than in lab schools or other idealized settings. Finally, they worked with teachers over a long period of time.

Still, despite this personalized, context-specific, and on-going assistance, most Follow Through sponsors would probably say that their models were only rarely implemented with the degree of fidelity that they wanted. Moreover, even the degree of fidelity they did obtain was not due entirely to their teacher education efforts, but also to a great deal of voluntary movement of teachers into and out of Follow Through classrooms. Over time, Follow Through classrooms came to be populated with teachers who were, prior to learning these models, already sympathetic with the general ideas the sponsors espoused.

The fact that the modest implementation that was achieved occurred only after such intense intervention and only after teachers self-selected themselves in and out of Follow Through classrooms, raises serious questions about the extent to which teachers can really learn different approaches to teaching. And it suggests

that wide-scale teacher education aimed at any particular model may be next to impossible. More than we did in the 1960s, we now realize that the task of teacher education entails more than teaching teachers specific techniques, and more than teaching them a vision of an ideal classroom. It entails both of these and more, for teachers must grasp the significance of these new ideas, must understand how these ideas differ from those they have held in the past, and must be persuaded that these ideas are better than the ideas they had in the past.

Several hypotheses can be put forward to account for lack of high-fidelity implementation of Follow Through models. One is that model sponsors did not really have fully developed models at the time they began teaching teachers, but instead had rough theories and visions of an ideal kind of classroom. Another is that they were not themselves experienced teacher educators, and did not know how to help teachers understand or adopt their nascent models. Probably few, if any of them, for instance, took into account the deeply-held and tacit convictions that teachers brought with them to Follow Through. If they did not, they may have tried to teach teachers how to behave without articulating fully their own assumptions about why this would be a superior way to behave. Or, if they did articulate their assumptions, they probably did not appreciate the magnitude of the task of persuading teachers than these ideas were superior to the ideas teachers already held.

Implications of the Models for Defining Children's Life Chances

I suggested in my introduction that the Follow Through models have implications both for teacher education and for whether and how we might do a follow-up study of Follow Through children. Though they went about it in many different ways, all Follow Through model sponsors expected their models to have a long-term effect—to benefit children long after they completed their formal participation in Follow Through. These benefits would occur either because children knew more to start with, and therefore would be better able to learn more, or because children would know more about how to learn on their own, or because children would have more supportive families.

And none of these theories, at least as I have broadly portrayed them, is idiosyncratic to an isolated Follow Through model sponsor. All of these paths to improved life chances are espoused by a substantial segment of our education community and by a substantial segment of our lay community as well. It is reasonable, then, to ask whether we can find any evidence that such long-term benefits did actually occur. And if they did, what these benefits look like now.

When the national evaluation of Follow Through was conducted, the third cohort was considered the test cohort. Students in this cohort were presumed to have participated in fully-developed models, and our most complete data exists for this cohort. Cohort 3 children entered kindergarten in the 1971-72 school year and completed third grade in 1975. They should have graduated from high school in 1984. They would be about 25 years old now, and would be around 27 years old by 1992, when a follow-up study might occur. They are clearly old enough now to make some claims about their life chances.

BEST COPY AVAILABLE

But if our interest is in life chances, we need not tie ourselves to this cohort. This is the cohort with the most *achievement* data, but we need not focus on achievement differences now. School achievement is, after all, only a proxy for other achievements in life. Since Follow Through graduates are now young adults, and have probably completed whatever education they aimed for, we can examine a wide range of other indicators of their life chances. I make this point only to suggest that, if sponsors believe the fourth, or even fifth cohorts might have been better implemented than the third, we could also follow-up on the children who were members of these cohorts. Even students in these cohorts would be in their early twenties by now.

But the most difficult issue we face in this follow-up study is not which cohort to follow; it is what to look for as indicators of improved life chances. I recommend two criteria for evaluating evidence of improved life chances. One is that the evidence be relevant to something we might define as "life chances." The second is that the evidence can be construed as related in some way to the efforts of Follow Through model sponsors. To appreciate the significance of these two criteria, consider some examples of evidence we might obtain from former Follow Through participants.

If we are interested in their life chances, one thing we surely should determine is whether they are still alive. Being alive is a necessary condition to life chances. But most of us would not consider this a sufficient outcome, largely because we would think it was not related in any way to the particular educational models we are examining. I nominate this indicator of improved life chances in part facetiously, but in part to illustrate my point: being alive is highly relevant to one's life chances, but not very relevant to an assessment of the Follow Through models. If we found differences in the death rates among participants of different models, we would have a difficult time explaining this difference in terms of the models.

But arguments could be made. One could argue, for instance, that attention to parents and community could have altered children's environments in ways that ultimately made them safer, more protective than they otherwise might have been. Perhaps those sites or models with most parental involvement would demonstrate greater life rates than others. Now suddenly the issue gets sticky, for an argument could probably be made to account for virtually any pattern of outcomes we find. Thus the problem of attaching meaning to our findings is a very serious one, and its salience in this proposed follow-up study is one reason why I insist on selecting evidence that is relevant both to life chances and to the ideas espoused by various Follow Through models. Let me consider each of these criteria separately.

Defining indicators of life chances

I am assuming that the kind of data we will probably collect falls into the category commonly called indicators. That is, we have no standardized measure of life chances, and once students are out of school, measures of achievement lose their cache. So we need to identify evidence that we believe indicates something about these young adults' current status and that indicates something about their life chances. For instance, indicators of life chances might include:

- Whether they are alive
- Whether they completed high school
- If not, whether they received a GED
- Whether they attended a community college
- Whether they attended a four-year college
- Whether they graduated from a community or four-year college
- Whether they are employed
- What income they earn
- Whether they are in jail

But even if we agree that most of these are reasonable indicators of life chances, we would probably all generate different models of how early childhood might have influenced these outcomes. On one hand, these outcomes seem so far removed from early childhood education that it would be difficult to develop a plausible argument that early childhood influenced any of them. On the other hand, the Follow Through models did aim to achieve some benefits of this kind, and the human imagination is remarkable in its ability to generate plausible scenarios for how almost anything could account for almost anything else. That is the main reason why I advocate indicators whose relationship to the models seems plausible at the outset, before data are collected.

Defining indicators of model effects

In addition to defining evidence that can be plausibly construed as indicating the life chances of young adults, we need to define evidence that can be plausibly related to the early childhood programs these people experiences nearly 20 years ago. This will be difficult because, although most of them claimed to be influencing life chances, they did not directly intervene to influence any of the outcomes I just listed. Instead, they influenced other events, which they hoped would, in turn, influence these events.

One solution to this problem, of course, would be to measure outcomes in these young adults that are more closely related to the outcomes that sponsors tried to directly influence. Since the models were, for the most part, intended to influence youngsters' educational achievements, we could define a number of education-specific indicators. I do not mean to suggest that we give these young adults yet another paper and pencil achievement test, or an intelligence test, but that we look at such things as:

- High school grade point averages
- Placement in remedial programs
- Placement in upper or lower track
- Achievement test scores on whatever tests they took while in school
- High school attendance

BEST COPY AVAILABLE

These indicators are more clearly related to education than those I listed above. Yet, despite their educational flavor, they still do not tie us very tightly to Follow Through models. We are still in a position wherein almost any pattern of outcomes could be interpreted in almost any way we wish. For, just as model sponsors did not directly influence such life events as employment or jail, they also did not directly influence high school curriculum choices.

Moreover, the closer we move toward those outcomes Follow Through models were designed to influence, the further we move from life chances. That is, it might be possible to try, once again, to actually measure all the outcomes models did try to alter: specific knowledge, problem-solving ability, self esteem, or perhaps even relationships with parents. But few people would take these as evidence, by themselves, of life chances. Most would be more persuaded of life chances by the kinds of evidence I listed earlier.

It should be clear now that the problem we face in defining evidence of life chances is closely connected with the problem we have drawing causal inferences from a study conducted some twenty years after our "treatments" were administered. The models themselves did not hold explicit ideas about exactly how these directly-influenced outcomes would eventually lead to the other outcomes that matter in life. In the absence of explicit arguments from the models themselves, we are free to generate whatever causal interpretations we want. The indicators, by themselves, will not help us with the general problem of attributing outcomes to Follow Through models.

The problem of drawing causal inferences from a follow-up study

The Follow Through Planned Variation Study was what researchers call a quasi-experiment. That is, even though it was not a perfect experiment, it was designed to draw on experimental logic to reach its conclusions. The logic of inference from experiment depends on statistical probabilities: We look to see whether children with one background, on average, differ from those with another background, and infer that the background differences are responsible for the other observed differences. If we were to extend this experimental logic to a 20-year follow-up, we might look to see whether young adults who had experienced one model were more likely to be gainfully employed, or less likely to be in prison, than those who participated in a different model, in the hope that we could infer that their early program experiences were at least in part responsible for these differences.

For a variety of reasons, it will be difficult to apply this logic to our hypothetical follow-up study. Though most people here know these reasons, I think it is important to enumerate at least three of them again.

1. Even when the study was first completed, when there was no time lag, and when the outcomes were more closely mapped to model goals than they would be now, there were tremendous difficulties attributing outcomes to models. Comparison groups were not really comparable, the settings in which different sponsors worked were not really comparable, and sponsors differed as much in their ability to develop materials and to train teachers as their models differed in their theories and visions of early childhood

education. These problems made even the initial third grade differences difficult to interpret, and will continue to make new data difficult to interpret.

2. Nearly 20 years have passed since these children finished the Follow Through portion of their education and there have been numerous other influences on their life chances during this long interval. These other influences could either enhance or detract from the initial goals of their model sponsors. That is, children who participated in a model that strove to encourage self-initiated learning could have later participated in a highly structured, direct instruction middle school program. Or vice versa. Even if we assume that most of the remainder of their educational lives was relatively ordinary, we cannot assume there were no extenuating circumstances outside of school. Implementation of the Follow Through models was inhibited by tornadoes, floods, collapsing buildings, teacher strikes, desegregation and several other events. No doubt these children have continued to encounter as many unusual events since then as they did in those four years. Just as life itself can be influenced by numerous events outside our control, so can most of these indicators of life chances be so influenced.

3. The attrition we encounter in this study will be so severe that it will confound virtually any statement we may want to make, for we will have virtually no idea of the extent to which our follow-up sample represents the original population of Follow Through participants. Add to this the fact that the fact that we will want to draw inferences about differences among models, not just about Follow Through in general, and that attrition rates may vary across the models. Differences in outcomes may be either masked or exaggerated, or both, by attrition.

The confounding influences of different implementation across sites and models, different intervening influences across students, and different attrition rates across sites and models are so numerous and complicated that causal inferences will be extremely difficult to draw. We could find ourselves once again wondering why we could not see clear model effects when the models themselves were so clearly different.

In fact, because of the capriciousness of life, we face a second difficulty as well: Just as it will be difficult to argue that any particular early experiences could have contributed to these outcomes we find, it will be difficult to argue that these outcomes are really indicative of future life chances. We already know, for instance, that although children of poor families are more likely to be poor than other children are, there is still a remarkable degree of income redistribution each generation. Even within a family, siblings may experience widely differing economic futures (Jencks, et al, 1972). Moreover, we know that differences in economic status are only weakly related to measures of academic achievement. And finally, we know that economic circumstances are not stable within a given adult. That is, there is considerable change within individuals over time (Bane and Ellwood, 1983; Duncan, 1985). I don't mean to suggest that economic indicators are the only criteria we should use to measure the life chances of these former Follow Through children, but only to suggest that a considerable amount of evidence has accumulated now suggesting that people's current status is not highly related to the two predictors we generally assume matter most: parental status and cognitive ability.

BEST COPY AVAILABLE

So identifying evidence that is both relevant to life chances and can be plausibly attributed to an early childhood educational experience is not a trivial matter.

An alternative approach

If we really want to draw plausible inferences about the impact of these early education experiences, we need to use a different research logic, perhaps replacing the logic of statistical probabilities with a clinical, or diagnostic argument. We need to rest our conclusions on an interpretation of *how* one event appears to influence another rather than on *how often* one event appears *in conjunction* with another. If we use a clinical or diagnostic argument, rather than an experimental argument, we also need to gather evidence that goes beyond indicators such as those I listed earlier and gather, in addition to these, some interpretations of how these young adults came to be in the situations they now are in. Why do they think they are well off, or not well off, for instance? Such interpretations could be offered either by the Follow Through graduates themselves, by their parents, or perhaps even by a former teacher who has stayed in touch. Through these interviews, we could learn such things as the following:

- What do they recall from their early education and what are their views are about the relevance of these experiences
- What are their views about how and why certain events happened to them. For instance, if they are employed, did someone else help them get the job, or did they get it on their own?
- Once in a job, do they keep it for long, do they lose it soon, or do they get promoted?
- Once in a college, do they work at their studies, drop out, fail, or succeed?
- To whom do they attribute their successes or failures—themselves, their families, coincidence, etc? If themselves, do they attribute their successes to their knowledge, perseverance, cunning, etc?

I would recommend interviewing both the Follow Through graduates and their parents, for they are likely to recall different things and to interpret the importance of events differently. Parents are more likely to recall the details of their children's early education experiences than their children will be, and may have their own interpretations of why events unfolded as they did. In those models that were designed to foster changes in parents, they may be more able to define ways in which these early encounters with schools altered future events in their lives and in their children's lives.

If we can interview participants and their families, and learn their own interpretations of events and how they perceive themselves, their past, and their futures, we may be able to generate plausible hypotheses about how these young adults came to be in the situations they are in. We may be able to estimate the extent to which their future life chances look good or bad, and the ways in which their early education might have contributed to their current status and to their current methods of interacting with their environments.

This form of evidence cannot, of course, save us entirely from the problem of sample bias. However, it will enable us to make reasonable arguments about those young people we are able to locate.

Conclusion

The arguments I have made here may seem overly pessimistic. With respect to the implications of the Follow Through experience for teacher education, you could interpret my position as suggested that it would be nearly impossible to prepare large numbers of teachers to adopt any of the Follow Through models. I have argued that it would be difficult to shake teachers' prior assumptions that teaching is a rather simple matter of telling what you know and to shake their assumptions that if students cannot learn from this form of teaching, the students are at fault. With respect to the potential of a follow-up study, you could interpret my position as suggesting that plausible inferences from a follow-up study are likely to elude us. I have argued that, even under the best of circumstances, we may find ourselves unable to interpret findings.

Still, the point of the program was to examine alternative hypotheses about how to improve children's life chances, and the evidence currently available on these models stops considerably short of informing us of this outcome. A follow-up study of young adults who participated in some of these models, regardless of its shortcomings, would give us more knowledge than we now have. Such a study need not be expansive, and need not canvass students from all models or all sites within a model. In fact, it may be more illuminating if it is more selective, focusing on locations where the models were better implemented to start with and where records of graduates are more complete.

References

Ball, D. L. (1988). *Knowledge and reasoning in mathematical pedagogy: Examining what prospective teachers bring to teacher education.* Unpublished doctoral dissertation, Michigan State University.

Bane, M. J. & Ellwood, D. (1983). *Slipping in and out of poverty: The dynamics of spells.* Working paper, Harvard University.

Duncan, G. (1985). *Years of Poverty, Years of Plenty.* Ann Arbor, MI: University of Michigan Press.

Feiman-Nemser, S., McDiarmid, G. W., Melnick, S. L. & Parker, M. (in press). Changing conceptions of beginning teacher education students: A description of an introductory course. *Journal of Education for Teaching.*

Floden, R. E. (in press). What teachers need to know about learners. In M. M. Kennedy (Ed.), *Teaching Academic Subjects to Diverse Learners.* New York: Teachers College Press.

Jencks, C., Smith, M., Ackland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effects of family and schooling in America.* New York: Harper.

Kennedy, M. M. (in press). Policy issues in teacher education. *Phi Delta Kappan.*

Lortie, D. C. (1975). *Schoolteacher: A sociological study.* Chicago: University of Chicago Press.

BEST COPY AVAILABLE

McDiarmid, G. W. (in press). Merging subject matter and diverse learners. In M. M. Kennedy, (Ed.), *Teaching academic subjects to diverse learners*. New York: Teachers College Press.

Paine, L. (in press). Teaching as a virtuoso performance: A Chinese model of teaching. *Teachers College Record*.

Slavin, R. E. (1988). Cooperative learning and student achievement. *Educational Leadership*, (October), 31-33.

Follow-Up of Follow Through: What Use for Policy?

Richard F. Elmore
Harvard University

The Problem

Follow Through was conceived in the waning years of the Great Society as an ambitious intervention both to improve schooling for disadvantaged children in the early grades of school and to increase our collective knowledge of what works educationally for these children. Although the program never reached the scale its advocates desired, it did, during its peak years of 1968-1978, account for about $500 million dollars in federal expenditures, of which about $30-50 million was for research and evaluation. (House, et al., 1978: 129) It was a unique event in the history of educational policy, involving as many as twenty research and development organizations that developed educational models, dozens of local school districts that implemented them, tens of thousands of children who were the recipients of services, a number of evaluation contractors, and hosts of kibitzers and critics. Within this frame of reference, it seems plausible to ask whether we could learn anything useful by following up the effects of Follow Through.

Follow Through is a very different kind of intervention from Head Start or Chapter 1, the federal government's other large-scale attempts at compensatory services for disadvantaged students. Head Start and Chapter 1 operate on the premise that the intervention is a combination of money and program guidelines which result in

services for participating children. A follow-up study of these programs, then, might simply compare participants with eligible non-participants on selected measures at various times in their later lives. Follow Through, on the other hand, operated on the premise that money and program guidelines were not enough. The additional element of the intervention was that each participant was to be exposed to a distinctive approach to classroom instruction. Follow Through was, in other words, not designed to be a single intervention, but a variety of different interventions, operating simultaneously under the rubric of "planned variation." Asking how Follow Through affected participants' later lives, compared with eligible non-participants, then, is not a particularly useful or interesting question, since the essence of Follow Through is planned variation. The more appropriate question is how children were affected by different approaches to compensatory education.

Hence, the essential problem is whether we can learn something useful from examining the effects of a number of different compensatory education programs on participants' later lives, or how these programs affected successive generations of participants. This additional requirement of examining different approaches makes both the design and the policy-relevance of a follow-up study much more complex than a study of single intervention.

Another problem associated with doing follow-up studies of earlier interventions is that the policy environment around the studies is never the same as it was when the initial intervention was undertaken. The issues, the key actors, the locus of policymaking activity, the fiscal and political constraints, and the potential audience that would be willing to pay for such a study and attend to its results all change over time. A follow-up study of Follow Through would be done in a dramatically different political world from the one in which Follow Through was initially conceived. The success of a follow-up study would be determined, in large part, by whether it could successfully connect with the current political environment.

This paper, then, focuses on the particular problems associated with doing a policy-relevant follow-up study of a planned variation intervention, and the problem of "fitting" a follow-up study to the current policy environment.

The Legacy of Follow Through

Most of the particular problems associated with doing a policy-relevant follow-up study of Follow Through stem from its history. It is commonplace to observe that Follow Through was, through its entire history, characterized by a tension between social action and social science. For many of its initiators and for virtually all its local constituents, Follow Through was a way of focusing extra resources on disadvantaged students; the objective of learning more about what works for disadvantaged students by studying planned variations in instructional programs was secondary. For the sponsors and evaluators, the planned variation objective was what justified the rather large expenditures over and above program services.

One form in which this tension manifested itself was in the scale of the program. The planned variation objective was grafted onto Follow Through as a way of increasing its appeal when its initial $120 million authorization was funded with a $15 million appropriation. A few program sponsors were recruited and sites were selected for a relatively small-scale test of alternative programs. As the program ap-

BEST COPY AVAILABLE

propriation grew, social action began to drive social science. More sponsors and sites were selected, without regard for their effect on the overall design of the planned variation component. At its peak Follow Through involved 20 sponsors, with each sponsor operating in a minimum of three or four sites. The planned variation objective required that every sponsor should be able to demonstrate its capacity to succeed. But the complexity of the design dictated by the number of sponsors and sites made any carefully-designed assessment of all sponsors virtually impossible. The evaluation eventually focused on 13 sponsors. In an attempt to rationalize the design, the evaluators grouped sponsors, for part of the analysis, into four main groups. This attempt to simplify the design and make it more intelligible, not surprisingly, raised strong objections from Follow Through sponsors and constituents, who felt that their distinctive approaches to education were being inappropriately lumped together for the convenience of a simplified evaluation design.

One important legacy of Follow Through, then, is that, at its peak, the program was simply too large and too complex to be usefully evaluated as a planned variation experiment. An important question that designers of a follow-up study will have to confront is how to simplify and focus the design. Simplifying and focusing, however, raises questions about whether it is Follow Through whose long-term effects are being studied or specific aspects of the intervention. If specific aspects, then which and why?

Another important legacy of the Follow Through program is that development—of instructional programs, of the evaluation design, of assessment instruments, and of the implementation of models in sites—consistently lagged behind the program's growth and the progress of the evaluation. Only a few of the initial sponsors actually had fully-developed programs and were prepared to implement them in sites when Follow Through began. As these sponsors developed their models and gained experience in the field, new sponsors were added, introducing new cycles of development. Some sponsors responded to the tasks of development and implementation systematically; others were less well-organized. Under the best of circumstances, with fully-developed models and a well-organized support system, there were large site-to-site variations within sponsors in implementation and effects. The overall evaluation design was in flux from about 1969 until 1975, when a design was imposed on the existing array of sponsors and sites in the interest of getting an evaluation completed. Program administrators and sponsors had always aspired to have outcome measures that reflected the true diversity of educational aims represented by instructional models. There was even talk of developing new outcome measures in order to meet this challenge. In the final analysis, evaluators used a battery of existing outcome measures that virtually all sponsors thought were extremely limited in their capacity to measure the educational effects of their programs.

Another legacy, then, is that major development issues in the program and the evaluation were never fully resolved, except by imposing a sort of artificial analytic order on programmatic disorder. This legacy raises a number of questions for designers of a follow-up study. Should a follow-up study attempt to establish some sort of continuity with outcome measures used in the original evaluation, or should it attempt to represent the existing state of the art in testing and measurement? Should a follow-up study revisit the issue of which measures are appropriate to specific models? And should a follow-up study attempt to take account of developmental and implementation problems that existed in the original program?

BEST COPY AVAILABLE

A third legacy, less tangible than the others but no less influential, is that the design of the Follow Through evaluation was probably not adequate to detect the kind of effects that the instructional programs could produce. As Robert Egbert has said on more than one occasion, the initiators of Follow Through believed in what he calls the "big bang" theory—that the effects of well-designed instructional programs on students' learning would be so dramatic that there would be no question about models' effectiveness. At the time Follow Through was initiated this might have been a plausible theory. Now it is much less so.

Judith Gueron, President of the Manpower Demonstration and Research Corporation (MDRC), an organization that has designed, evaluated, and assisted in the management of a number of randomized field experiments in welfare and employment over the past fifteen years, puts this problem succinctly. She argues that no social intervention, particularly one directed at chronically disadvantaged clients, can show large initial effects. MDRC's work, however, clearly demonstrates that even small initial effects can have large social consequences. For example, a small difference favoring treatment over control participants in a program to prepare young unemployed adults to enter the labor force can produce a relatively large benefit-cost ratio as the earnings of participants accumulate over time. (For an analysis of experimental versus non-experimental studies of employment training programs see, Fraker and Maynard, 1987; Barnow, 1987; and Burtless and Orr, 1986.) Gueron's point, which was made several times in the course of the Follow Through evaluation with less evidence to back it, is that we should expect small initial effects from interventions and design studies to detect them. Hence, MDRC specializes in designing randomized field experiments in which operating agencies of government administer the interventions in "real world" settings. Several of these experiments have shown small effects with impressive social pay-offs.

This third legacy—Follow Through's weak design for small effects—suggests that any follow-up study should embody an explicit theory, or a set of theories, about how initial effects become long-term effects. Also, a follow-up study should honestly acknowledge in its own design that the design of the original evaluation probably wasn't powerful enough to pick up effects of the magnitude one should expect from an intervention of that type. One major decision in the selection of a theory, which I will address in the final section of this paper, is whether the study should trace the effects of Follow Through on individual students as they advance through school or whether it should focus on the effects of the intervention operating through schools on successive cohorts of children.

The Follow Through Evaluation as Policy Research

For an effort that involved a relatively large proportional expenditure on research, development, and evaluation, Follow Through has had a notably modest impact on policy. The program probably had significant benefits for local participants, sponsors, and evaluators, most of which were unmeasured and unreported by the formal evaluation. There is little evidence, however, that the evaluation had much effect on the content of policy or the design of programs for the educationally disadvantaged. After the brief and modest flurry that greeted the release of the report of the national evaluation in 1977, not much has been heard about Follow Through as policy research.

Some of the reasons for this lack of attention to Follow Through as policy research stem from institutional factors. Much of the argument over the effects of Follow Through was probably seen by policymakers as special pleading by the program's constituents, rather than as serious discourse about how policies toward disadvantaged students should be designed. Furthermore, in the late 1970s and early 1980s, the country's domestic policy agenda shifted dramatically away from designing more effective federal policies to limiting federal expenditures and shifting authority for the initiation of policies away from the federal level.

Some of the reasons for Follow Through's lack of visibility as policy research are more substantive, however. Not the least of these is that the program-wide results of Follow Through, at least as measured by the national evaluation, showed, at best, negligible effects on the learning of disadvantaged students. The three main findings of the national evaluation, as reported by Abt Associates, the evaluation contractor, were:

1. that "differences in outcomes between sites within each model were greater than overall differences in effectiveness between models;"
2. that "Follow Through groups scored about as one would expect similar disadvantaged groups to score without Follow Through," [and] "where differences were apparent, Follow Through groups scored lower more frequently than they scored higher;" and
3. that "with few exceptions, Follow Through groups were still scoring substantially below grade level at the end of three or four years' intervention" (Anderson, et al., 1978: 162-163). A re-analysis of the national evaluation data showed similar results (House, et al. 1978: 149-150).

The one major positive finding from the national evaluation, which was widely reported at the time the report was released, was that so-called "basic skills" models seemed to have more positive effects on students' learning and self-concept than so-called "cognitive-conceptual" or "affective-cognitive" models (Stebbins, et al., 1977). A re-analysis of these results, however, seemed to show that most of the positive effects on student learning resulted from one model's performance on the math and language subtests of the Metropolitan Achievement Test (House, et al., 1978: 149-150). Based on these findings, the national evaluation concluded that "externally sponsored curricular change is not a reliable tool for raising the test scores of poor children" (Anderson, et al. 1978: 162).

Rejoinders to these findings focused on the perceived failure of the national evaluation to capture the real meaning and effects of the program and on the host of methodological and analytic decisions that went into the evaluation. Representatives of the program's funding agency, the U. S. Office of Education, stressed the positive side of the finding that site-to-site differences exceeded model-to-model differences. "Compensatory education can work," they argued. "There were enough instances of success in enough sites to suggest that the Follow Through children did better than they would have done in the absence of the program" (Wisler, et al. 1978: 179). As noted above, the decision to group models into broad categories came under heavy attack on the grounds that it had little or no basis in theory or practice. Sponsor representatives stressed the tangible benefits that the program had delivered to schools and children. Sponsors demonstrated that new approaches to educating disadvantaged children could be implemented and that

teachers could change deeply entrenched patterns of practice. Parent engagement in education seemed to increase in many sites. And the program resulted in the development of an impressive array of materials and a new level of engagement between research and practice (Hodges, 1978).

One can argue, of course, that large-scale attempts to integrate policy, social science, and educational practice, no matter how sophisticated they are methodologically, will always generate high levels of disagreement over results. Nonetheless, the negative tone of the national evaluation report and the muddiness of the debate that followed it raise serious questions about the political feasibility of a follow-up study. Why would it be useful to study the long-term effects of a program that demonstrated largely negligible short-term effects? Given the importance of site-to-site variation in determining program impacts, would a follow-up study really be a study of Follow Through's impact, or a study of the impact of some particular set of local adaptations in response to a federal initiative? How would we expect the results of a follow-up study to differ—substantively and methodologically—from the original evaluation? Would a follow-up study raise all the same chronic methodological and analytic questions that plagued the original evaluation? Who would be the audience for such a follow-up study, given that there is no demonstrated need to know—particularly at the federal level—more about the kind issues that the original Follow Through evaluation raised?

It seems highly implausible, then, to propose a simple extension of the original Follow Through evaluation as a follow-up study. There are other options for framing a follow-up study, but these options should be developed with some knowledge of the changes that have taken place in the political environment since the late 1970s.

The Shifting Environment of Education Policy

Since this subject is a digression from the main topic of the paper, let me summarize as succinctly as possible, in a few blunt assertions, how I think the political environment around education policy has changed since the late 1970s. I would like to use these assertions as a way of framing some alternative ways to think about a Follow Through follow-up study.

First, the locus of policy initiative has shifted from federal to the state and local levels. The Follow Through evaluation was published at the end of a period of maximum federal influence and initiative in education policy. The federal role in education has declined substantially since the late 1970s. At the same time, there has been a dramatic upsurge in state and local policymaking activity. In the early 1980s, this activity focused mainly on state mandates designed to increase graduation requirements, curriculum requirements, and to increase the quality and performance of teachers. These policies produced predictable amounts of local- and school-level variation in implementation and outcomes, but, for the most part, did not result in discernible impacts on student outcomes. In the last four of five years, the focus of state and local policy has shifted to measures designed to increase the capacity of schools to teach high quality academic content to children from a wide variety of backgrounds. This shift has meant more attention at the state and local level to the problems of restructuring schools around more challenging conceptions of teaching and learning, and restructuring relationships among states, localities, and schools to reflect more capacity-building and less direct regulation. This shift

has also meant that a number of states are beginning to redesign state policies on curriculum, testing, and teacher preparation and professional development to organize them around more coherent, simpler, systemic views of what students should learn. (See Fuhrman, Clune, Elmore, 1988; Firestone, Fuhrman, Kirst, 1989.)

Given the political and fiscal constraints under which the federal government is currently operating, it seems unlikely that the federal government will re-emerge as the leading initiator of policy. At the same time, there is no sign that the level of state and local policy initiative will decrease. Any ideas for new policy research initiatives connected with past federal programs, then, should take account of the fact that the main audience for these initiatives will be state and local, not federal, policymakers.

Second, discourse about the content of education policy has shifted markedly from remediation and equity for selected groups of students to improving the quality of education, and increasing expectations, for all students. Our conception of what constitutes an adequate education has focused more sharply on ambitious conceptions of learning, using such terms as "high literacy," "higher order thinking," and "teaching for understanding." Lauren Resnick captures this shift when she argues, "The goals of increasing thinking and reasoning ability are old ones for educators...Although it is not new to include thinking, problem solving, and reasoning in *someone's* school curriculum, it is new to include it in *everyone's* curriculum" (Resnick, 1987: 7). One important component of this line of argument is an explicit criticism of the implicit assumption embodied in earlier research, policy, and practice that mastery of so-called "basic skills" precedes "higher order" skills. The new conventional wisdom is that this relationship is reciprocal; one needs certain basic knowledge of facts and algorithms to master higher-order, inferential knowledge, but the acquisition of basic facts and algorithms depends on the acquisition of higher skills (Ibid.). The distinctions that served as the basis for many of the conclusions in the Follow Through evaluation—that there is a difference, for example, between so-called "basic skills" and "cognitive-conceptual" approaches—is clearly out of synch with much current thinking about the nature of teaching and learning.

Third, beginning in the early 1980s with the so-called "effective schools" movement and extending through the efforts of such current reformers as Theodore Sizer and James Comer, there has been a shift in our conception of the relevant unit to which policy interventions should be addressed. From the mid-1960s to the late-1970s, policies were largely designed to target resources on specific groups of students within schools. From the early 1980s to the present, we have increasingly thought of policies as designed to improve the functioning of *schools*, including their responsiveness to children with differing educational needs. A large part of current policy discourse, then, is organized around proposals to improve the capacity, organizational health, and performance of schools, rather than to remediate the needs of specific groups of students within schools. Follow Through straddles this distinction. While for purposes of logistical convenience and philosophical conviction, Follow Through sponsors tended to view schools as the unit of intervention, the evaluation clearly focused on students as the unit of intervention.

Fourth, and closely related, there is a developing disenchantment among both policymakers and policy analysts with categorical programs that target specific groups of students for special attention. Among the criticisms of categorical

programs is that they tend to isolate students from the mainstream experience of schools, they balkanize and increase the complexity of districts and schools by introducing specialists and administrative requirements, and this balkanization leads to inefficient allocation of resources for all children's learning. Much of current policy discourse, then, is organized around ways of breaking the existing categorical structure that particularly characterizes urban schools with large proportions of disadvantaged students and finding new ways to organize instruction that focus on diverse student needs in a simpler more efficient structure (Elmore & McLaughlin, 1988). Again, Follow Through seems to straddle this distinction. Follow Through advocates thought of the program as a comprehensive approach to compensatory education, designed to address the broad educational, social, and emotional needs of children, but the program and the evaluation had a decided categorical focus on children who met certain family income requirements.

A follow-up study that adopts the same basic questions as the earlier evaluation, that assumes an audience made up largely of federal policymakers, that follows the distinctions drawn in the original evaluation between basic and higher order cognitive skills, that treats individual children rather than schools as the unit of intervention and analysis, and that adopts an essentially categorical model of policy will probably not be a very useful study. One could argue, of course, that the design of the original program and evaluation constrains what one can do with a follow-up study, hence, if the initial design doesn't fit well with the existing policy environment, it doesn't make sense to do a follow-up study. It is conceivable, however, that a follow-up study might be designed to have a tighter fit with the current environment. I will speak to this issue in a moment.

Research is more likely to influence policy and practice when it addresses questions that are being asked by policymakers and practitioners. These questions are, in large part, a function of where policy discourse is centered—who the key actors are, what the dominant terms of discourse are, where existing research and policy initiatives are headed, and what models of intervention appear to be most promising. Simply extending an earlier study into the future, while it may have intrinsic appeal for the basic questions it answers, isn't likely to have much influence on policy.

Some Options for a Policy-Relevant Follow-Up Study

The central issue, then, as I have posed it, is how to design a follow-up study of Follow Through that operates within the constraints set by the original program, that doesn't simply repeat the same tiresome methodological and analytic issues raised by the earlier national evaluation, and that, in some way, engages the interest of policymakers around current questions. This is a tall order. I'm not sure it's possible, but it is certainly worthy of some systematic consideration. I will sketch out three possible options, with the caveat that they probably don't begin to exhaust the full array of options available but they could simulate some useful thinking. I will label each option by the units of analysis that would form the point of departure for its design, and I will attempt to characterize the kind of study that would flow from a focus on a particular point of departure. I will also attempt to discuss some of the practical problems of constructing treatment and comparison groups within each option.

Selected sponsors, selected sites

The most obvious option, and the least attractive in terms of my criteria, is to try to replicate some version of the 1977 national evaluation—estimating the effects Follow Through models on participants over time. This option would, as the original evaluation did, focus on sponsors, using individual students, nested in sites, as the units of analysis. The theory behind this study, as with the national evaluation, is that Follow Through had its primary impact on individual students through the programs delivered by sponsors.

Such a study could take at least two forms. One would be simply to fund a few existing sponsors to enhance their data collection and analysis of longer-term effects on students, consistent with their earlier work, and perhaps to construct longitudinal data bases in sites where they previously hadn't been able to do so. My hunch is that the design of a follow-up study would take a pragmatic turn at some point and focus on the handful of sponsors that have sustained a substantial research capacity independent of Follow Through over time and that have, on their own constructed, longitudinal data bases. Two logical candidates would be, for example, High Scope and the University of Oregon. This form of study would, or course, raise the question of whether sponsors could provide an adequately objective account of their own performance; during the national evaluation, sponsors' evidence on program effects consistently provided a rosier account than the evaluators'. Another form of this kind of study would be to fund an independent contractor to reanalyze existing data on the longitudinal effects of selected sponsors, or possibly to collect data in sites where such data hadn't been previously collected. This latter form would, of course, raise all the old political issues about whether external evaluators could really understand what sponsors were trying to achieve.

In either case, comparison groups would probably have to be constructed from the records of eligible non-Follow Through students within sites. To the extent that sponsors have already constructed such groups in their earlier research, the issue is evaluating the adequacy of these groups for further longitudinal study. To the extent that sponsors haven't constructed these groups, the issues are more complex. Any follow-up study that relies on data not previously collected by sponsors will be quickly driven to existing record data on attainment and achievement within school districts, again raising questions about the appropriateness of the outcome measures for individual sponsors. Such data varies widely in content and quality by school district and is quite expensive to collect. If the object is to follow-up on the national evaluation, samples will have to be constructed on students who started elementary school in 1970 or 1971, who probably exited elementary school in 1977 or 1978, and who graduated from high school in 1983 or 1984. The period of 1970 to 1984 is one of dramatically varying capacities in school districts for testing and record-keeping. A follow-up study would tend to gravitate toward sites with relatively sophisticated testing and record-keeping capabilities; there is no way of knowing what proportion of original Follow Through sites in certain models have this capacity without surveying them.

Student attrition is an enormous problem. School districts typically keep basic records on attainment as long as students attend district schools, but their records on where students go after they leave the system are either non-existent or very poor, especially during the period in which Follow Through students from the

national evaluation would have been in school. Annual turnover rates in excess of fifty percent are not unusual in urban schools.

The point of such a study, as policy research, would be to say something useful about the question of whether early intervention improves students' chances for later success in school. It is highly implausible to think of doing such a study on the scale of the original evaluation. So a key problem in designing the study would be how to reduce the scale of the original to focus on a small collection of sponsors and sites that would yield potentially worthwhile results. Given the largely negative results of the original evaluation, it seems to me that a threshold condition for such a study would be some evidence, either from the national evaluation or from well-designed research by the sponsors, that the sponsor had some discernible positive effect on students at the beginning of the period. Studying the longer-term consequences of early non-effects doesn't seem to me to be particularly attractive investment opportunity for scarce educational research dollars.

One can easily imagine the cascading sampling decisions that would drive a follow-up study based on the national evaluation: sponsors who could demonstrate positive effects, sites with promising record data, students on whom adequate record data are available, etc. It is questionable what one would actually know after such a study were done. Such a study would probably be more an extended essay on the difficulties of doing longitudinal research on originally messy data than a study of the long-term effects of early intervention.

Schools, sponsors

Another possible option would be to break the lock of the original design somewhat and to treat Follow Through as an intervention that was designed to change the way *schools* treat disadvantaged students. In this case the unit of analysis would be schools, crossed by sponsors, rather than individual students nested in sites. The theory behind this study is that Follow Through had its primary impact on students by altering the way schools operate, hence the important question is not how the program affected participants over time but how the program affected the way schools treat successive cohorts of children with diverse needs.

This design would treat schools as relatively stable institutions through which cohorts of students and interventions travel in regular cycles. At some point in the life history of these organizations called schools, Follow Through entered their structure and certain things happened as a result of this intervention. At some point, presumably, Follow Through left the structure. Schools respond differently to these external interventions and their organizational responses have different effects on students. Rather than assuming that schools exist to implement Follow Through, as the original evaluation and the first option do, this design assumes that schools are places where teachers teach and students learn in presence of a variety of influences, including external interventions like Follow Through.

Such a study would select a sample of schools from among those that sponsors worked with, presumably from a reduced list of sponsors and presumably with some thought to the length of sponsor involvement and the sponsors' perceived success of that involvement. The design might vary the length of sponsors' involvement and sponsors' evaluations of the schools' degree of model implementation.

The study would focus on how schools deal with cohorts of students, many of whom carry the characteristics traditionally associated with educational disadvantage, over a specified period of time. So instead of attempting to follow students who participated in Follow Through into their later schooling, the study would focus on how schools responded to successive cohorts of children from diverse backgrounds over some period that included active involvement with Follow Through sponsors. Also, this design puts no special premium on the "success" of Follow Through sponsors in the national evaluation, although such a criterion could be introduced. It would presumably be of some value to policymakers to know what happened to the schools after the withdrawal of Follow Through support; so the design should include a substantial proportion of schools whose participation in Follow Through was active for a period of time and then ceased. It would also be of some value to policymakers to know whether interventions like Follow Through improve the capacity of schools to educate children over time; so the design should account for schools with varying periods of participation.

Part of the study would be a qualitative analysis of the life-cycle of innovations in schools, focusing on the way the school handled student diversity and the role that Follow Through played in shaping the school's response at various stages of its involvement and disengagement. One would presumably want to look for traces of Follow Through's effects on the present structure of schools. Another part of the study would be a quantitative study of patterns of student composition and achievement over some specified period of time. One could think of this as a before, during, and after longitudinal design where data were available, or simply as a during and after design.

While such a study is still heavily reliant on local data sources, it is considerably less sensitive to the vagaries of local data collection than the first option. Many more systems collect school-level achievement data, for example, than reliably collect data that can be traced to individual students over time. So focusing on the school as the unit of analysis relieves many of the problems connected with resurrecting bad individual-level data from school district archives. Such a study will probably not produce comparable data on student achievement across schools in different districts over time. However, it is possible to think of clusters of schools with different attributes that happen to have used similar achievement tests over time. Since the relevant unit of comparison is the school and not the individual student, problems of constructing comparison groups become more tractable.

The main advantage of this design, however, is not methodological convenience. Its utility has more to do with its policy-relevance. The design could say something useful about the effects of Follow Through on schools and students, and about the relationship between the intervention and patterns of school-level achievement over time. The main utility of this design, however, would be what it would contribute to our understanding of how *schools* respond to diverse student populations in the face of changing external conditions. This issue is of considerably greater long-run significance, with all due respect, than a study of how well Follow Through worked as a planned variation intervention. We might actually learn something from such a study about how to design more effective external interventions by understanding how schools integrate, or fail to integrate, these disturbances into their existing institutional structures.

BEST COPY AVAILABLE

Teachers, schools, sponsors

This option is an elaboration of the previous one to include teachers and teaching practice as an explicit factor in determining what students are exposed to. The theory behind this study is that Follow Through had its primary effect on students by altering the way teachers teach, but that influences on teaching are first filtered through the organizational structure of schools.

Just as schools vary in the way they deal with external interventions, so too do teachers within schools. One could imagine a design like the previous option that also included relatively detailed data on teachers' involvement in Follow Through and other attempts to influence their instructional practice, possibly crossed with student achievement data.

The logic behind this design is the same as the previous one at the next lowest level of aggregation. Teachers form small organizations for learning called classrooms within larger organizations called schools. They too deal with problems of student diversity within the context of a variety of supports and constraints, occasionally including helpful (or not so helpful) people from the outside who want to change their teaching practice. It would be interesting to know how teachers respond to various attempts to influence their practice, including interventions like Follow Through, and whether these interventions have any residual effects after they are withdrawn.

The design of this option would be essentially the same as the previous option—schools crossed by sponsors, with data on cohorts of students--but would include some attempt to map teachers' participation in Follow Through in addition to school-level data. This design adds considerable complexity to data collection, since it assumes some capacity to disaggregate student data for any given cohort to the classroom level. This capacity may not exist with enough frequency in local districts to make the design practical. In the absence student achievement data disaggregated to the teacher level, however, it would still be useful to have a series of profiles of teachers who worked in Follow Through and in its absence to understand the relationship between external interventions and teaching practice.

Conclusion

Follow Through represented a certain period of thinking about how to mobilize social science research in the service of social policymaking. The present policy environment is quite different from the period in which Follow Through was conceived and implemented. A follow-up study of Follow Through that attempts to revive the original questions posed by the Follow Through evaluation will also revive all the methodological, analytical, and substantive issues associated with the original evaluation. Since the policy agenda has moved on in many respects, it seems pointless and not very productive to revisit these issues.

It is conceivable, however, that a follow-up study could be designed that would speak to important contemporary questions. For all its defects, Follow Through was a concerted effort to change the way schools work and teachers teach by bringing to bear external resources in the form of outside experts. We could learn a

considerable amount from this experience that might be useful in designing current state and local policies. In order for this to happen, however, the focus of the study has to shift away from studying how individual students respond to an intervention and toward studying how an intervention influences teachers' practice and student learning in schools over time.

References

Anderson, R., et al. (1978). "Pardon Us, But What Was the Question Again?: A Response to the Critique of the Follow Through Evaluation." *Harvard Educational Review*, Vol. 48, 161-170.

Barnow, B. (1987). "The Impact of CETA Programs on Earnings: A Review of the Literature." *Journal of Human Resources*, Vol. 22, 158-192.

Burtless, G. & Orr, L. (1986). "Are Classical Experiments Needed for Manpower Policy?" *Journal of Human Resources*, Vol. 21, 606-638.

Elmore, R. & McLaughlin, M. (1988). *Steady Work: Policy, Practice, and the Reform of American Education*. Santa Monica, CA: The Rand Corporation.

Firestone, W., Fuhrman, S. & Kirst, M. (1989). *The Progress of Reform: An Appraisal of State Education Initiatives*. New Brunswick, NJ: Center for Policy Research in Education, Rutgers University.

Fuhrman, S., Clune, W. & Elmore, R. (1988). "Research on Educational Reform: Lesson on the Implementation of Policy." *Teachers College Record*, Vol. 90, 237-257.

Fraker, T. & Maynard, R. (1987). "The Adequacy of Comparison Groups for Evaluations of Employment-Related Programs." *Journal of Human Resources*, Vol. 22, 195-224.

Hodges, W. (1978). "The Worth of the Follow Through Experience." *Harvard Educational Review*, Vol. 48, 186-192.

House, E., et al. (1978). "No Simple Answer: Critique of the Follow Through Evaluation." *Harvard Educational Review*, Vol. 48, 128-160.

Resnick, L. (1987). *Education and Learning to Think*. Washington, DC: National Academy Press.

Stebbins, L., et al. (1977). *Education As Experimentation: A Planned Variation Model, Volume IV-A, An Evaluation of Follow Through*. Cambridge, MA: Abt Associates.

Wisler, C., et al. (1978). "Follow Through Redux: A Response to the Critique by House, Glass, McLean and Walker." *Harvard Educational Review*, Vol. 48, 171-185.

Potential Benefits of Longitudinal Studies in Education

Herbert J. Walberg
University of Illinois at Chicago

To gain a broad perspective on the possible benefits of longitudinal studies and how they might apply to Follow Through, this paper draws together some findings from studies of the influences of education and psychological environments on learning and development. My intention is to provide some factual background for a synoptic view of learning, human development, and accomplishments over the life span. Since other papers in this set focus on the limited number of investigations of early childhood effects, I will concentrate on the more common school-age through adult studies and try to draw some lessons for policy and future longitudinal studies.

It seems worth saying at the outset that no single study or method can be viewed as definitive. If many well-designed and executed studies point in the same direction, however, then a reasonable amount of confidence can be placed in their conclusions. But for many psychologists, the term "well-designed studies" means experiments with randomized assignment of subjects to treatment and control groups to eliminate causal uncertainties. Few long-term educational studies meet this criterion. Experiments, however, are hardly fool-proof since they usually fail the sociological criterion of broad generalizability. Nearly all experiments are restricted to idiosyncratic rather than random national samples; and they often employ better

92

executed treatments than are found in ordinary practice. Nor are one-shot surveys panaceas. Inducing causality from cross-sectional surveys is suspect since a conclusion that "A causes B" may ignore reverse causation and third causes of both A and B.

Longitudinal studies can offer distinct advantages: Some causal uncertainty may be eliminated by agreeing that later events do not cause earlier events, and that later psychological traits and states do not cause earlier ones. Valid and comprehensive measures of earlier and later personal attributes and intervening conditions may allow rough estimates of causal influences. Still, designing such analyses requires difficult choices such as large samples vs. intensive measurements, and random vs. program samples. Despite great effort and expense, moreover, subjects, especially from highly mobile impoverished populations, will be lost in increasingly larger numbers over long time periods.

The lesson to be gained is that only those findings that emerge strongly and repeatedly from multiple studies employing different methods can be trusted. Syntheses of previous research, theory building, and secondary analyses of existing data can all help in this effort. Longitudinal studies, however, do have a special place: Other things being equal, they afford a better opportunity than cross-sectional studies to probe causality, and a better chance than most experiments to generalize widely. Carried out over long periods, moreover, they reveal not only immediate effects such as learning but long-range adult outcomes such as occupational success, civic participation, and avocational accomplishments.

Some examples

The benefits of longitudinal studies may be best seen in specific findings. They may be striking enough to change personal practices, re-think interventions, and challenge stereotypes. Consider three recent illustrations.

Some 60 million Americans suffer from hypertension; and millions are treated with drugs to diminish their high blood pressures despite undesirable side effects. A "meta-analysis" (statistical summary) of nine studies of 43,000 moderately hypertensive (diastolic readings of 90 to 105) patients followed over an average of 5.6 years showed that drugs did not significantly reduce coronary heart disease mortality (Ornish, 1990, p. 62). Without such persuasive longitudinal research, symptoms might continue to be treated. As a consequence, the quality of life would have been diminished without reducing mortality. The search for efficacious treatments, moreover, might have been discontinued.

Many social analysts assume that the poor families live in big Northern cities and engender generation after generation of poverty. But longitudinal research belies this stereotype. Duncan's (1984) longitudinal analysis of about 5,000 American families showed substantial social mobility. For example, of families in the top or bottom 20 percent of income in 1971, only about half remained in these classifications in 1978. Between 1969 and 1978, 25 percent of the families fell below official poverty lines in at least one year, but less than 3 percent remained below in 8 of the 10 years. Even these persistently poor belied stereotypes: two-thirds lived in the South; one-third were elderly; and only a fifth lived in large cities.

Education was a minor influence in determining changes in wealth or poverty. Most decisive were voluntary changes in family structure—marriage or divorce, a birth or a child leaving home. Job-related changes such as layoffs and physical disabilities were second in importance.

Another remarkable example is Werner and Smith's book *Vulnerable but Invincible*, a 30-year study of 698 at-risk infants born on the Hawaiian island of Kauai. It shows how some children triumph over physical disadvantages and deprived childhoods. Despite prenatal and perinatal stress, discordant and impoverished home lives, uneducated, alcoholic, addicted, and mentally disturbed parents, some went on to develop healthy personalities, stable careers and strong interpersonal relations. Some were even hospitalized and separated from their families for extended periods; many as infants and toddlers had mothers who worked full time and had no access to stable child care. Some were babies of single or teen-age parents with no other adult in household; others were migrant and refugee children without permanent roots in community.

Resilient children were protected by consistent nurturing which encouraged trust in its availability. At least one person in their lives buffered stress by accepting them unconditionally, regardless of temperamental idiosyncrasies or physical or mental handicaps. Grandparents, older siblings, and teachers provided such bonds in cases in which parents were incapacitated or unavailable. Such informal ties to kin and community were more powerful than government-provided services.

The Value of Learning

Given these notable examples, we can turn to a more searching history of longitudinal effects of educational programs. My purpose is to compare the quantitative magnitudes of effects of educative factors not only on immediate academic learning but on real-world outcomes in adult life. For several reasons, however, it seems best to concentrate on learning in the first several decades of life. Much deliberate learning takes place at this period in the life course; language, habits, tastes, job skills, and much else are largely formed by age 30; and such early investments in learning pay dividends over a long period. Considerable research on learning in this period has been conducted and quantitatively synthesized (Walberg, 1984b). Learning processes in later adult life, moreover, can be hypothesized to be similar in principle to those of the formative years.

First, it is important to recognize that much what we know about the causes and effects of learning is restricted to a great extent to what can be measured on academic and non-academic tests. These tests may be comprised of oral, essay, or multiple-choice questions; but their common feature is the measurement of knowledge, skills, and the like through words. Thus, investigations of causes and effects of education and other stimulating experience are often confined to verbal achievement. Perhaps this preoccupation is short-sighted, but a reasonable case can be made that the primary medium of education, work, and life is language. Those who do well at it can be expected to be successful at the many tasks that require language competence over the life span.

Bormuth (1978) concluded that verbal activities consume much time and are of immense importance in adult work and leisure. In a 1971 survey of about five-thousand people aged 16 and over, 87 percent of those gainfully employed reported that they had to read as part of their jobs; and typical working people read for 141 minutes per day as part of their jobs, or about 29 percent of the work day. Since the national wage bill in 1971 was $859 billion in 1971, Bormuth estimated that U. S.. workers earned $253 billion for on-the-job reading. In addition, considerable leisure activity is spent on reading. Of the total sample, for example, 73 percent reported having read a newspaper the previous day; the mean reading time for this activity was 33 minutes. Thus, the estimated time devoted to reading and its imputed value are huge.

How demanding is the typical reading task? Diehl and Mikulecky's (1980) case studies assessed the literacy requirements of a broad range of 100 occupations and high schools. They interviewed and tested 107 adults ranging from a lawyer and other professionals to assembly-line operatives and stonecutters in a sample of 26 workplaces within a 70-mile radius of Bloomington, Indiana, including Indianapolis.

They found that general on-the-job reading is sometimes repetitive; many reading materials can be seen as continuous reminders rather than demanding verbal tasks. On-the-job technical reading, however, is often demanding, even more demanding than reading in high schools, as shown by other surveys reviewed by the authors. Technical workers encountered a wider variety of reading materials, read them more competently and comprehended them in greater depth than did the students. Professional workers read an even greater range of materials and read them more competently than technical workers and students. All groups of workers, moreover, saw reading as more important to their success than did the students.

Adult development

Aside from verbal demands alone, Kohn and Schooler (1983) showed that intellectually stimulating work contributes to cognitive development in adult life. Follow-up surveys conducted as recently as 1973 of a 687-respondent sub-sample first tested in 1964 showed that intellectually and socially demanding jobs fostered continuing growth in intellectual functioning. Other surveys show that adult motivation and language activity foster test achievement (Frederick and others, 1980), although all three are predicted by stimulating family and educational experience in childhood (Walberg & Tsai, 1983).

What can be measured on verbal tests is neither unimportant nor a complete indication of educational accomplishment and preparation for adult life. Such tests indicate little about the perseverance, creativity, integrity, social sensitivity, team skills, and other learning goals that educators, parents, and students in English and U. S. surveys consistently rank above verbal achievement as such (Raven, 1981).

Extracting information from verbal media, however, may constitute a quarter or third of adult work and probably now amount to over $500 billion dollars per year

in compensated adult time. If verbal skills fostered in school such as writing, listening, and speaking, as well as literacy activities in childhood and adult leisure are added, then their value is immense, and the time they take may be the biggest slice of waking life. Improvements in verbal-educational skills by increasing educational productivity in schools, and leisure and work settings may be very much in the individual and national interest.

In addition, there is little evidence to show that improving verbal test achievement sacrifices other valuable traits. Evidence cited below suggests that, to the extent verbal and other symbol-manipulative skills are more deeply and widely learned in educational and occupational experience, national welfare and economic productivity may be increased. On the other hand, verbal skills hardly guarantee integrity, social sensitivity, character, parental expertise, and on-the-job competence although they may contribute to such desirable adult characteristics.

It seems ironic that schooling and other educative experiences, which constitute such a large fraction of human time and which may have immense consumption and investment value, are so narrowly and poorly measured. Much more information is available on automobiles; for example, base price, accessory and maintenance costs; information on speed, safety, size, and reliability; and ratings on styling and handling. Comparatively little information of this kind is available on the costs and benefits of education.

Educational Productivity

A vast literature of more than 8,000 psychological studies of the past half century contains estimates of the quantitative correlations and effects of the proximate factors that bear directly on student achievement. During the past five years these studies have been quantitatively synthesized (see Walberg, 1984b, for detailed description). The estimates show that nine psychological factors produce much larger and more consistent effects than class size, staff salaries, expenditures per student, and other crude indicators of quality.

The nine factors appear to require optimization to increase affective, behavioral, and cognitive learning. Potent, consistent, and widely generalizable, these factors fall into three groups:

Student aptitude includes:

- ability or prior achievement as measured by the usual standardized tests,
- development as indexed by chronological age or stage,
- motivation or self-concept as indicated by personality tests or the student's willingness to persevere intensively on learning tasks.

Instruction includes:

- the amount of time students spend in learning and
- the quality of the instructional experience, including psychological and curricular aspects.

BEST COPY AVAILABLE

Four environmental factors also consistently affect learning, the educationally-stimulating, psychological climates of:

- the home,
- classroom social group, and
- the peer group outside school; and
- minimum leisure-time television viewing.

The first five aspects of student aptitude and instruction are prominent in the educational models of Benjamin S. Bloom, Jerome Bruner, John B. Carroll, Robert Glaser, and others (see Walberg, 1984b, for a comparative analysis); each appears necessary for learning in school; without at least a small amount of each, the student may learn little. Large amounts of instruction and high degrees of ability, for example, may count for little if students are unmotivated or instruction is unsuitable.

These five essential factors, however, are only partly alterable by educators since, for example, the curriculum in terms of lengths of time devoted to various subjects and activities is somewhat determined by diverse economic, political, and social forces. Ability and motivation, moreover, are influenced by parents, by prior learning, and by students' contributions. Thus, educators are unlikely to raise achievement substantially by just their own efforts.

The psychological climate of the classroom group; enduring affection and academic stimulation in the home; an out-of-school peer group with learning interests, goals, and activities—these influence learning in two ways: Students learn from them directly. These factors, moreover, indirectly benefit learning by raising student ability, motivation, and responsiveness to instruction. In addition, about ten (not the more typical 30) weekly hours of television viewing seem optimal for learning, perhaps because more television time displaces homework and other educationally and developmentally constructive activities outside school.

The major causal influences flow from aptitude, instruction, and psychological environment to learning, although these factors also influence one another. Early achievement appears to raise not only the stock but the rate of learning. Called "the Matthew effect" after the "rich-getting-richer" passage of the Bible, the phenomenon appears fairly pervasive; children who start well at academic work and other endeavors gain at a faster rate and thereby gain increasingly larger advantages as they grow older. It seems that early success may increase motivation and also attract parental and teaching attention to the possibility of developing high talent and accomplishment. Complex reciprocal causation or mutual enhancement of ability, motivation, instruction, and stimulating environments over the early life course probably account for the Matthew effect (Walberg & Tsai, 1983).

The first five essential factors appear to substitute, compensate, or trade-off for one another at diminishing rates of return. Immense quantities of time, for example, may be needed to bring about a moderate amount of learning if motivation, ability, or instructional quality is minimal. Thus, no single essential factor overwhelms the others; all appear important.

Quantitative syntheses of thousands of experimental and quasi- experimental studies suggest that these generalizable factors are the chief influences on cognitive, affective, and behavioral learning (Walberg, 1984). Many of these studies provide

strong causal inferences since they are generally true experiments with random assignment or quasi-experiments with pretests to measure longitudinal gains so as to equate treatment and control groups.

The productive factors, however, were also probed for their significance in promoting learning in large sets of survey data on elementary and high school students—the National Assessment of Educational Progress, High School and Beyond, the Scholastic Achievement Tests, School Health Educational Evaluation, and the Study of Mathematically Precocious Youth. Also included were several smaller, localized surveys. The 23 studies of about 250,000 students in six subjects of study contained 341 regression weights, 303 or 88.8 percent of which were signed as hypothesized (all positive except amount of leisure-time television viewing).

Many of these analyses, however, were cross-sectional and are subject to causal ambiguity. New analyses, however, of the Longitudinal Study of American Youth show that, controlled for one-year earlier pretests, science and mathematics achievement and attitudes are causally influenced by the nine-factors in the directions indicated by theory and previous research. Similar results can be expected from analyses of the National Elementary Longitudinal Study.

Education and Adult Accomplishments

It may be assumed that the first four to six years of schooling are necessary to acquire literacy, numeracy, and some basic knowledge and skills needed to function in modern society. Beyond this, of what use is further education? The usual view is that additional education promotes acquisition of knowledge and skills useful in the pursuit of subsequent education, work, and leisure. In accord with this view, it is clear that people with more education do better in these respects, although it is less clear that education is the cause of their success.

Real-world accomplishments

Another view is that the usual academic courses of classroom lectures, discussions, homework, and the like merely promote the acquisition of academic facts and concepts, which are measured and graded on the usual verbal-educational tests, but which have little to do with other present and future accomplishments in "the real world" (Wallach, 1970). This view is consistent with the statistical facts that measured intelligence predicts grades and standardized achievement scores; but none of these have much to do with the student's other accomplishments nor adult success.

In a large, varied group of working scientists, for example, Harmon (1963) found no association of indexes of professional accomplishments such as patents, publications, and prizes with their verbal and mathematical aptitude tests, achievement tests in their fields of concentration, and grades in science courses. Neither did Bloom (1963) nor Helson and Crutchfield (1970) find associations of IQ and other mental abilities with distinguished accomplishments, as rated by knowledgeable peers in samples of chemists and mathematicians. Hoyt (1966) reviewed studies

BEST COPY AVAILABLE

relating grades and later occupational success in business, engineering, medicine, and scientific research, and found no association. In Great Britain, Hudson (1960) similarly found no association of academic and scientific distinction; and the same lack of association has been found in samples of architects, artists, writers, and other occupations (see, for example, Taylor & Baron, 1963).

Research on high school students shows the same absence of association. Their grades and test scores predict college grades, but neither predict non-classroom accomplishments in the humanities and sciences at either level. Non-classroom accomplishments during the high school years, however, predicted with a modest degree of accuracy similar accomplishments in college and in adult life after the college years (Richards, Holland, & Lutz, 1967; Munday & Davis, 1974).

In an Educational Testing Service report for the College Entrance Examination Board, Breland (1981) questioned reviews of such research and defended the use of standardized tests for college admission. He summarized 275 studies from colleges; business, graduate, law, and medical schools; and business, industrial, governmental, and medical settings that related prior characteristics to learning outcomes such as grade-point averages, degree attainment, and performance ratings in leadership, science, and musical, dramatic, literary and graphic arts.

Prior indicators including recommendations, interviews, and measures of interest and personality were nearly useless in predicting either academic or non-academic accomplishments (among 8 median correlations the highest was .28 which accounted for less than 8 percent of the variance). "Biodata" or responses to questions about life experiences predicted academic and nonacademic outcomes (.43 and .35); but they are impractical for competitive academic selection since they are easily faked. The only substantial prediction was the familiar academic outcomes from prior grades and test scores (.60 which on average accounted for 36 percent of the variance).

Granting that asking a testing agency about the value of tests is like asking a barber if you need a haircut, it seems incontrovertible that tests and prior grades predict subsequent grades with moderate accuracy. Grades, however, have little or no value in predicting subsequent success.

Grades in higher education

Grades are one indication of educational accomplishment. Surveys of grades and adult outcomes, though restricted to higher education, provide little evidence that good students by this index turn out much better than their peers. In 35 studies of business, military, and civil-service people, as well as teachers, professors, scientists, physicians, and nurses, grades accounted for 2.4 percent of the variance in income, self-rated happiness, and job satisfaction, numbers of patents and publications, and effectiveness ratings by peers and supervisors (Samson and others, 1984).

All these groups, of course, are more homogeneous than the general population in abilities indexed by grades that presumably make for personal and career success; such small variance limits covariance and therefore predictive validity. Having a

degree, moreover, even though one's grades were poor, probably ingures greater special competence and leads to somewhat higher income on average than that of a comparable person without such a degree. Still, the low correlations inspire little confidence that added effort to achieve higher grades results in greater competence than that gained by specific training, on-the-job experience, and character traits.

Perhaps employers may be quite rational in assigning only moderate importance to grades in making decisions. Surveys of employers indicate that, beyond some fairly minimal academic competencies, responsibility, social skills, and the like are heavily weighed (Walberg & Sigler, 1975).

Crain's (1984) recent national survey of 1,283 recruitment and employment personnel showed that 94 percent rated dependability as "extremely important" in hiring high school graduates; 82 percent rated proper attitudes about work and supervisors similarly, and 74 percent rated teamwork also as "extremely important." Smaller percentages gave ratings of extreme importance to cognitive skills such as: rapid learning, 57 percent; reading materials as difficult as a daily newspaper, 56 percent; reading complex material, 22 percent; and handling complex calculations, 11 percent.

In Plato's triumvirate of affect, cognition, and behavior, hearts—for some purposes—may deserve even more attention than minds and hands, at least more than the academy seems willing to credit. Yet, at the other end of the cognitive-skill spectrum, limited verbal skill can be a substantial handicap to an individual, corporation, or nation. The Conference Board's (1984) survey of about 500 human-resource and public-affairs executives showed that most agree that many newly hired high school graduates have difficulty in reading and understanding instructions and in expressing themselves clearly to co-workers and supervisors.

Occupational tests

Hunter and Hunter (1984) found, moreover, that cognitive knowledge and skills are the best predictors of job success. They reviewed studies of efficiency induced by using cognitive ability tests to select employees. Such tests increase efficiency by allocating the most able applicants to occupational positions. High selection validity; that is, predictability of job performance from test scores, makes for financial savings in public and private organizations. Hunter and Hunter estimated, for example, that if the Philadelphia Police Department were to drop the use of its cognitive selection test for entry-level officers and choose them at random, it would cost the city government $170 million over a decade.

For jobs in the federal government, they estimated that the substitution of criteria other than cognitive tests for entry-level hiring would cost from $3.1 billion (job tryout) to $15.9 billion (age) in lost productivity per year. They further estimated that the productivity differences between avoidance and complete use of cognitive tests for the U. S. workforce as a whole for 1980 would amount to at least $80 billion, equal to total corporate profits for the year.

Such estimates are based on performance criteria such as supervisor ratings, work samples, and production records. Hunter and Hunter's compilation of validity coefficients document the impressive record of cognitive paper-and-pencil

tests. Their analyses showed that ability composite tests, which can be given in as little as a half hour by untrained clerks, are by far the most valid for entry-level prediction. Interviews, on which much reliance is placed, are far lower in validity, and combining them with ability tests would only raise the validity at most from .53 to .55. For current employees, moreover, ability tests predict supervisor ratings nearly as well as the best on-the-job assessments of present performance.

Economic Returns to Education

In his widely noted book, *The Over-educated American*, Freeman (1976) estimated that the social rates of short-term returns for a college education declined from the range of 11 to 14 percent from 1950 through 1970 to about 7.5 to 9.5 percent in the 1970s. Because of a possible surplus of degreed young people, he speculated that the decline in the relative rate of return to college graduates since the mid-1970s may persist.

McMahon and Wagner (1982), however, argued that longer-term rates of return provide more valid estimates because the labor market needs time to adjust to changes in educational standards, and because the superior returns of more education may not show up for a decade or so beyond graduation. Using long-term rates of return, they showed no decline in the relative market value of a high school or college education since 1958. In fact, from 1958 through 1976, the social returns to a college education, about 12 to 16 percent, have been superior and more stable than returns of stocks listed on the New York Stock Exchange (which centered on zero to three percent and were often negative during the period).

McMahon and Wagner further showed that, despite the higher opportunity costs of foregone earnings of those admitted to select professions, private rates of return are highest in medicine, law, and engineering-technical fields. They are lowest in the clergy, natural science, social science, and education fields.

Analysis of social returns to education in other countries provides a basis of comparison with U. S.. results. Psacharopoulos's (1983) analyzed social returns to education in 44 developing, intermediate, and advanced countries. The average returns were all positive and range from about 10 percent (required for World Bank investment loans for physical-capital investments) to considerably higher rates.

These educational investments in human capital, however, like many other increases in factors of production, show diminishing returns. Primary education in developing countries pays the highest returns, and higher education in advanced countries pays the least. As Psacharopoulos points out, education in primary schools and in elementary textbook subjects is cheaper to provide and consumes less valuable human capital than advanced, technical education.

All such monetary-returns comparisons may underestimate the full benefit of education because they omit its consumption value, for example, joy in learning, as well as the later psychological satisfaction with work it may confer (Lucas, 1977). They may also overestimate the complete benefit by failing to calculate time and pleasure given up in youth to pursue a select profession and also given up in later family life and friendships to pursue demanding careers. As the next section indicates, however, such subjective factors and non-monetary returns are exceedingly difficult to estimate.

Non-Monetary Returns

Michaels (1982) points out that education may raise "productivity in non-labor-market activities and thereby provides non-monetary benefits" (p. 141). "Yet here, as elsewhere in this research area, one has the impression that a rather low burden of proof regarding a non-market productivity effect is imposed" (p. 131). In other words, as in much non-experimental research, the burden of proof lies with opinion or "theory" rather than varied, stringent tests of hypotheses. Some results and uncertainties, however, are worth considering.

Education appears to increase knowledge of personal asset management and propensity to save. The more highly educated adopt new products more quickly; more highly educated women are more likely to use contraception and have fewer unplanned children. A high school education apparently raised a scaled index of married men's health by 3.5 percent (holding age constant) and 1.2 percent holding constant prior health status, socioeconomic status, measured ability, income, obesity, and wife's schooling. Each additional year of higher education raises the probability of excellent health by an estimated one percent.

Yet, as Michaels points out, these small associations may be accounted for not by the direct effects of schooling but by the rational use of the more valuable (income-imputed) time of the more highly educated. Because their time is more costly, "the more educated have an incentive to be relatively well informed. So whether it involves videotape machines, vasectomies, or credit cards, the finding that the more educated adopt a new product more readily is not surprising, nor does it necessarily reflect a differential productivity effect of schooling. Higher income which accompanies schooling may determine the observed behavior" (p. 133).

Other possible benefits of education are successful marriages and families, and in social ascendancy. Going to school may be a good investment in finding a desirable spouse. It has been estimated, for example, that a wife's schooling raises her husband's earnings by roughly half as much as does his own schooling. Schooling and higher education, moreover, may foster private inter-generational returns by assortative mating by intelligence (Becker, 1981); and more highly educated families are repeatedly found to migrate longer distances when they move, which may lead to better job opportunities.

These findings are consistent with the hypothesis that education increases knowledge of opportunities. But they are also consistent, as Michaels points out, with other hypotheses: In addition to education, heredity, early family life, and parental socioeconomic status there may be prior or mutual determinants of education, knowledge, and opportunities; and the causal directions and weightings remain controversial.

Long-term effects of education

One of the few longitudinal studies that yields stringent estimates of the long-term effects of education, with social background and early intelligence statistically controlled, was conducted in Sweden. In 1961, Harnqvist (1984) surveyed about 12,000, or 10 percent, of the Swedes born in 1948, and followed up random samples of them with mail and interview surveys in 1980. Education, as compared with parental social class and measured intelligence, showed pervasive effects on

adult characteristics of men and women. Those with greater amounts of education liked school more than others; but even early school leavers saw the need for more education for the present generation of children. Those with more education more often described their job as providing opportunities for new knowledge, and they reported they had more influence on their own working conditions than did others.

Cultural activities such as reading literature and going to theaters and concerts were more frequent among the more educated. By contrast, entertainment activities such as reading weekly magazines, and viewing television and sports events were more frequent among the less educated.

The more educated had more frequent contacts with co-workers; the less educated more often saw family and relatives. Highly educated men reported higher skills in cooking and lower skills in car repair; having better information about how to appeal decisions and less about seeking economic support from society. In one-hour interviews, the more educated men used more words and proportionately more different and longer words. To finish the interviews, the less educated required far more prompts and interventions to complete the answers. Such pervasive education effects on life activities and verbal competence seem likely to be found in other countries.

The Terman study

The most famous longitudinal study in American psychology was begun by Terman (1925) and is still continuing (Tomlinson-Keasey & Little, 1990). The Terman data on gifted children spans a sizable fraction of the twentieth century. The study followed a group of bright American children periodically assessing personality characteristics, social attitudes, and mental capabilities to ascertain changes due to developmental or environmental factors. An advantage of this study is that IQ is almost controlled (participants were limited to those with IQs over 135) thus making other influences such as home environment and personality variables more apparent.

For his study Terman defined the gifted child, as one whose IQ score is 135, equalled by only about one in two hundred of the school population. The age range of the 643 grade school students was approximately four to fourteen years with a median age between nine and ten. Later 378 high school students were added to the study. Some birthdays were as early as 1908 to 1918. Today survivors could be 72 to 82 years of age.

The testing protocol included the Stanford-Binet Intelligence tests, Stanford Achievement Tests, a general information test, tests of interests, and medical examinations. Field workers made home visits for parent interviews and assessment of the home environment.

In 1928, students and parents again were interviewed and questionnaires completed. Later in 1936, 1950, 1955, 1960, 1972 and 1986 questionnaires were mailed to the students. As the students married their spouses were included to a limited degree.

The farsightedness of Terman is seen in his choice of variables. His thoughtfully prepared study includes measures of personality factors, social behaviors, instructional factors, as well as measures of home environment such as the number of books in the home, parental marital happiness, and the amount of time spent reading.

Gifted children were found to have more hobbies than other children. Terman suggests that the close relationship between the IQ and reading scores is, "probably because gifted children are such omnivorous readers." Gifted girls were found to score higher on masculine tests than the average girl and were less interested in girlish type activities. Gifted children were found to score above average on measures of emotional stability (Woodworth-Cady test).

To obtain a clearer view of factors that might influence the career achievement of the gifted another dimension was added to the study in 1940. The data of the upper 25%, the "A" group, and the lower 25%, the "C" group was compared in an effort to study the impact of factors other than intelligence on adult outcome. Terman defined success as, " the extent to which a subject had made use of this superior intellectual ability."

Using this approach Terman found a difference in parent educational levels. Twice as many "A" parents as "C" parents were college graduates. A fathers were more likely to be professional. Divorce or separation was more numerous among "C" families than "A" families. There was no difference in academic achievement during the elementary school years but a clear difference in academic achievement emerged during high school. Ninety-seven percent of "A"s entered college and 90 percent of them graduated while 68 percent of the "C" group enrolled in college and 37 percent graduated. Terman reports that data collected from the case histories and trait ratings by parents and teachers in 1922 further differentiated the "A" and "C" groups in adulthood. These observations are consistent with Freud's assertion of the long term effect of childhood experiences. Terman's findings are supported by studies of early life experiences of eminent men (Cox, 1926; Walberg, 1981).

A modern multivariate analysis also showed significant relations of early traits and conditions to later adult accomplishments (Tomlinson-Keasey & Little, 1990). Nonetheless, the sizes of the relations were very small. None of the early personality traits correlated higher than .14 with adult adjustment and occupational achievement. Intellectual skill correlated weakly and positively with occupational achievement but negatively with adult personal adjustment. Early family harmony, however, correlated about .3 with adult personal adjustment.

Socioeconomic status and other attainments

A vast literature bears upon the inter-generational question of parents' socioeconomic status (SES) on their children's academic development and life success. Many ambitious surveys and secondary analyses by economists and sociologists such as Featherman, Griliches, Hauser, Jencks, Lazear, Taubman, and others bear upon this question (see recent reviews by the economist Michaels, 1982, and the sociologist Sewell, 1981). Other studies have examined the influences of parents' education on the cognitive development and health status of their children.

Current reviews conclude that student learning is consistently associated with levels of parent education. But it may prove difficult to synthesize quantitatively the unique effects of parental SES in the context of rival causes as revealed in multivariate analyses because the constructs and their measures vary widely among investigations, thereby reducing the comparability of the causal weightings. These differences are attributable to differences in operationalization and in *a priori* theory (taken implicitly as a set of untested assumptions rather than hypotheses) on which investigators have achieved little consensus.

The SES influences on academic learning, moreover, may operate in several ways that are difficult to separate: Taking, for example, parental education as an index of SES, knowledge of childrearing, as well as higher income—conferred in part by more parental education—may both contribute to parenting capacity and effects on children's learning. More highly educated parents, for example, may themselves provide superior direct services to their children as a result of their knowledge; but they may also purchase, as a consequence of their higher income, superior childrearing goods and services.

One question, however, that can now be answered concerns the simple association of parental SES and children's learning. White (1976) collected 636 correlation coefficients of parental education, occupation, and income indexes of socioeconomic status (SES) with ability and academic achievement from 101 published and unpublished studies. In White's synthesis, the average correlation of learning with parent income, occupation level, and education are respectively 0.31, 0.20, and 0.19. (The correlations may be over- or under-estimated because of restriction of range in sampling, unreliability of measures, and other reasons.)

That income correlates the highest of the three may suggest that wealthy parents may confer more decisive advantages than highly educated parents by: purchasing time-saving household goods such as dishwashers and thereby being able to spend more time with their children, buying intellectually-enhancing toys and books for them; or hiring parent-surrogate services, such as day care and tutors, to nurture them. Little should be made of these speculations, however, since the SES correlations are all small and differ only slightly from one another.

Both the mean and standard deviation of all SES-learning correlations are 0.25; so that, on average, SES accounts for six percent of the variance in learning. Thus, contrary to great importance given to parental SES by some educational sociologists, its association with learning is surprisingly weak; and SES may constrain learning and social ascendancy far less than many believe.

Education and social mobility

If education by itself does not decisively influence adult success as measured by various indexes, neither do SES and other aspects of social background. Walberg and Weinstein (1984) analyzed the statistical dependencies of adult outcomes on 25 indexes of social background (including age and sex of the respondent and parental characteristics), diplomas and degrees, and a vocabulary test obtained on about two thousand men and women in the General Social Survey.

All independent variables in combination accounted for only small amounts of estimated variance in adult outcomes ranging from 3 percent of the variance in

happiness to 43 percent in occupational prestige. Family background, diplomas and degrees, and verbal competence together accounted for less than 13 percent of the variance in self-reported income, health, and happiness. Among the combinations of predictors and outcomes, diplomas and degrees uniquely accounted for the largest amount of variance in occupational prestige; but this amount was a trivial 2.3 percent.

Education and happiness

These findings seem typical of recent associations of educational and other measured formative effects on various adult outcomes. Only one of these, however, has been quantitatively synthesized. Witter, Okun, Stock, and Haring's (1984) collection of 176 zero-order correlations of self-rated well-being from 90 studies showed that amount and quality of education accounted for only about 1 to 3 percent of the variance in indexes of life satisfaction and happiness. When the association was controlled for occupational prestige, the variance estimates were even smaller; and the association has apparently remained constant for the past half century.

It appears at best that social background, education, and verbal competence in combination give adults slight to moderate advantages on indicators of adult success. Their separate influences, however, are weak—perhaps nil—and difficult to detect. Although they remain systematically and statistically undocumented, many other factors, such as accidental opportunities and personal initiative and perseverance, may play far larger roles.

Conclusion

This selective review of longitudinal studies in education and related research shows that we can be reasonably confident about what determines school learning. Syntheses of results of experiments and quasi-experiments as well as cross-sectional and longitudinal surveys all point at nine factors that consistently appear to influence learning. These factors may suggest the kinds of variables that are promising for early childhood longitudinal research on what makes for success in school.

We can be far less sure, however, about how these and other determinants in the first two decades influence adult accomplishments, partly because success in education hardly guarantees success in careers and other adult endeavors. More studies, better conceived and executed, might illuminate many pressing questions. They are well worth doing; but, by definition, they take time, and, are, as the adage has it, easier said than done.

References

Averch, H. A., Carroll, S. J., Donaldson, T. S., Kiesling, H. J., & Pincus, J. (1972). *How effective is schooling? A critical review and synthesis of research findings.* Santa Monica, CA: Rand Corp., Prepared for the President's Commission on School Finance.

Becker, G. (1981). *A treatise on the family*. Cambridge, MA: Harvard University Press.

Borger, J., Oh, S., Low, C., & Walberg, H. J. (1984). *Effective schools: A review of reviews*. Chicago: University of Illinois Office of Evaluation Research.

Bormuth, J. R. (1978). Value and volume of literacy. *Visible Language*, 12, 118-61.

Coleman, J. S., Campbell, E., Hobson, C, McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC: U. S. Government Printing Office.

Conference Board (1984). *The role of business in pre-college education*. New York, NY: Conference Board Research Bulletin RB 160.

Crain, R. L. (1984). *The quality of American high school graduates: What personnel officers say and do about it*. Baltimore, MD: Johns Hopkins Center for Social Organization of Schools.

Diehl, W. A., & Mikulecky, L. (1981). The nature of reading at work. *Journal of Reading*, 24, 227-247.

Duncan, G. J. (1984). *Years of poverty, years of plenty*. Ann Arbor, MI: Institute for Social Research.

Economist, Fast fries or high tech? June 16, 1984, 291, n. 7346.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37, 15-23.

Frederick, W. C., & Walberg (1980). Learning as a function of time. *Journal of Educational Research*, 73, 183-94.

Freeman, R. (1976). *The Over-educated American*. New York, NY: Academic Press.

Gagne, R. M. (1970). Policy implications and future research. In *Do teachers make a difference?* Washington, DC: U. S. Government Printing Office.

Glasman, N. S., & Biniaminov, I. (1981). Input-output analyses of schools. *Review of Educational Research*, 51, 509-539.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management*, 1, 19-41.

Harnqvist, K. (1984). *An empirical study of long-term effects of education*. Tel Aviv, Israel: First International Conference on Education in the '90s.

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary- school quality on academic achievement across twenty-nine high- and low-income countries. *The American Journal of Sociology*, 88, 1162-1194.

Horn, E., & Walberg, H. J. (1984). Achievement and interest as functions of quantity and quality of instruction, *Journal of Educational Research*, 77, 227-232.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.

Kohn, M. L., & Schooler, C. L. (1983). *Work and personality: An inquiry into the impact of social stratification.* Norwood, NJ: Ablex.

Levin, H., & Rumberger, R. (1983). The low-skill future of high tech. *Technology Review*, 86 (6), 18-22.

Lucas, R. E. B. (1977). Hedonic wage equations and psychic wages in the returns to schooling. *American Economic Review*, 67, 549-58.

Marshall, T. O., & Lohnes, P. R. (1965). Redundancy in student records. *American Educational Research Journal*, 2, 19-23.

McMahon, W. W., & Wagner, A. P. (1982). The monetary returns to education as partial social efficiency criteria. In W. W. McMahon & T. G. Geske (Eds.), *Financing education.* Urbana, IL: University of Illinois Press.

Michaels, R. T. (1982). Measuring non-monetary benefits of education: A survey. In W. W. McMahon & T. G. Geske (Eds.), *Financing education.* Urbana, IL: University of Illinois Press.

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. *Reading Research Quarterly*, 12, 400-19.

Ornish, D. (1990). *Program for reversing heart disease.* New York, NY: Random House.

Postlethwaite, T. N. (1975). The surveys of the International Association for the Evaluation of Educational Achievement: Implications of the IEA surveys of achievement. In A. C. Purves & D. U. Levine (Eds.), *Educational policy and international assessment.* Berkeley, CA: McCutchan.

Paschal, R. A. & Starhia, W. E. (in press). Educational productivity studies: A quantitative synthesis. In H. C. Waxman (Ed.), *Study of learning environments monographs: Volume 5.* Perth, Australia: Curtin University Centre on Science Education.

Psacharpoulos, G. (1983). *Education and National Development.* Washington, DC: World Bank.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *Elementary School Journal*, 83, 427-453.

Raven, J. (1981). The most important problem in education is to come to term with values. *Oxford Review of Education*, 7 (3), 253-272.

Riley, M. W., Hess, B. B., & Bond, K. (1983). *Aging in society: Selected reviews of recent research.* Hillsdale, NJ.: Lawrence Erlbaum.

Samson, G. E., Graue, M. E., Weinstein, T., & Walberg, H. J. (1984). Academic and occupational performance: A quantitative synthesis. *American Educational Research Journal*, 21, 311- 21.

Scott, R., & Walberg, H. J. (1979). Schools alone are insufficient. *Educational Leadership*, 37, 24-25.

Schultz, T. W. (1963). *The economic value of education.* New York, NY: Columbia University Press.

BEST COPY AVAILABLE

Schultz, T. W. (1981). *Investing in people*. Berkeley, CA: University of California Press.

Simon, H. A. (1983). *Models of bounded rationality: Vol. 2*. Cambridge, MA: MIT Press.

Sewell, W. H. (1981). Notes on educational, occupational, and economic achievement in American society. *Phi Delta Kappan*, 77, 322-325.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69, 213-235.

Terman, L. M. (1925). *Genetic studies of genius: Mental and physical traits of a thousand gifted children, Vol. 1*. Stanford, CA: Stanford University Press.

Tomlinson-Keasey, C. & Little, T. D. (1990). Predicting educational attainment, occupational achievement, intellectual skill, and personal adjustment among gifted men and women. *Journal of Educational Psychology*, 82, 442-445.

Tyack, D. B. (1972). The one best system: A historical analysis. In H. J. Walberg (Ed.), *Rethinking urban education*. San Francisco, CA: Jossey-Bass.

Walberg, H. J. (1974). *Evaluating educational performance*. Berkeley, CA: McCutchan.

Walberg, H. J. (1984a). Improving the productivity of America's schools. *Educational Leadership*, 41 (3), 19-26.

Walberg, H. J. (1984b, May). *National abilities and economic growth*. Chicago, IL: University of Illinois Office of Evaluation Research; presented at the American Association for the Advancement of Science, New York.

Walberg, H. J. (1973). Scientific literacy and economic productivity in international perspective. *Daedalus*, 112 (2), 1-28.

Walberg, H. J., & Shanahan, T. (1983). High school effects on individual students. *Educational Researcher*, 7, 4-9.

Walberg, H. J. & Sigler, J. (1975). Business views education in Chicago. *Phi Delta Kappan*, 56, 610-612.

Walberg, H. J., & Tsai, S. L. (1983). Matthew effects in education. *American Educational Research Journal*, 20, 359- 374.

Walberg, H. J., & Weinstein, T. (1984). Adult outcomes of connections, certification, and verbal competence. *Journal of Educational Research*, 77, 207-212.

Walker, D. F. (1976). Toward comprehension of curricular realities. In L. S. Shulman (Ed.), *Review of research in education*. Itasca, IL: F. E. Peacock Publishers.

Walker, D. F., & Schaffarzick, J. (1974). Comparing curricula. *Review of Educational Research*, 44, 83-111.

Weinstein, T., Boulanger, F. D., & Walberg, H. J. (1982). Science curriculum effects in high school: A quantitative synthesis. *Journal of Research in Science Teaching*, 19, 511-522.

Weiss, A. (1984). Simple truths of Japanese manufacturing. *Harvard Business Review*, 62 (4), 119-125.

Werner, E. E. & Smith, R. S. (1982). *Vulnerable but invincible*. New York, NY: McGraw-Hill.

White, K. R. (1976). *The relationship between socioeconomic status and academic achievement*. Unpublished dissertation, University of Colorado.

Witter, R. A., Okun, M. A., Stock, W. A. & Haring, M. J. (1984). Education and subjective well being: A meta-analysis. *Educational Evaluation and Policy Analysis*, 6, 165-73.

BEST COPY AVAILABLE

Strategies for Subject Identification, Location, and Interviewing

Sherri Oden
High/Scope Educational Research Foundation

The major purpose of this paper is to propose strategies for identification, location, and interviewing of former Follow Through program staff, students, and parents approximately 18 years since the students completed Follow Through as third graders. The recommended strategies are primarily based on research experiences in several major long-term follow-up studies conducted at High/Scope Educational Research Foundation. In particular, I draw upon experience from a current study, High/Scope's *Head Start Long-term Study*, a 20-year follow-up study of former Head Start students, now young adults. A high proportion of the model Head Start students also were in Follow Through classrooms in kindergarten. As I summarize this research, I discuss the implications for a national Follow Through follow-up study and propose specific methodological issues, approaches, and strategies that should be taken into account in the planning and implemintation of a Follow Through study.

In planning and carrying out the Head Start long-term study, now being completed, a number of published studies of other long-term studies have been particularly helpful: a forty-year follow-up study of middle aged adults who had been identified as mentally retarded (Ross, Begab, Dondis, Giampiccolo, & Meyers, 1985); the long-term follow-up of the *Family Development Research Program* from Syracuse

University (Lally, Mangione, & Honig, 1988); the ten-year follow-up of the *Family Support Intervention* (Seitz, Rosenbaum, & Apfel, 1985), and the research from the *Consortium for Longitudinal Studies* (Lazar, 1982). Many issues and strategies outlined in a variety of publications from the University of Michigan's Institute for Social Research (ISR) have also been helpful (e.g. Lansing & Morgan, 1971; Guenzel, Berckmans, & Cannell, 1983).

The Head Study Longitudinal Study

Throughout the last twenty years, the High/Scope Educational Research Foundation has conducted a number of longitudinal studies of graduates of various educational programs. Certainly the most well-known project is the *Perry Preschool Study*, i.e. *Changed Lives* (Berrueta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984), which is a longitudinal study of disadvantaged preschoolers who either participated in the High/Scope preschool program, or were in a randomly assigned control group. These subjects were studied in their elementary school years and at ages 15 and 19 years of age. We have just concluded a follow-up of these study subjects at age 28 years. The preschool curriculum model used in that research was subsequently employed in the national *Head Start Planned Variation Study* and was the basis for High/Scope's participation in national Follow Through programs in children's elementary school years.

Two of High/Scope's model sites from the *Head Start Planned Variation Study* found early IQ gains for children from its Head Start classrooms, compared to regular Head Start counterparts (Smith, 1975). The two sites are Ft. Walton Beach, Florida, with its regular Head Start comparison in Pensacola, Florida and Greeley, Colorado, which had both model and comparison Head Start classrooms. As indicated previously, a high proportion of these students were in Follow Through classrooms in kindergarten through third grade. Implementation ratings from the national observational data collection found that the model preschool classrooms (i.e., High/Scope) in these two sites were well implemented. For these reasons, we returned to these sites to conduct a 20-year follow-up study of High/Scope and regular Head Start graduates. We also constructed a no-preschool comparison group for this follow-up study. Each site included former students who participated in both Head Start and Follow Through and some students who participated in either one or the other of these programs. This research was funded by Health and Human Services (ACYF, Head Start Bureau) and the Spencer Foundation.

Program sites and groups

Ft. Walton Beach, Florida, is a small, tourist-oriented town with a pocket of poverty neighborhoods with mainly low-income black families. Also included in this site is Crestview, a neighboring rural area with both low- income black and white families. Both sites are a part of the Okaloosa County School District, and the High/Scope Follow Through model was implemented in two of its elementary school classrooms. In this site, only the Ft. Walton Beach subjects were in both Head Start in preschool and Follow Through in elementary school. Even in 1970, Pensacola was a more urban area than Ft. Walton Beach, but it was then selected to serve as the regular Head Start comparison group. The Head Start students

from Pensacola were also largely from black low-income families. Cantonment, a rural area just outside Pensacola, also had a regular Head Start program that was coordinated by Pensacola Head Start. Students from this area were from both white and black low-income families. Pensacola and Cantonment are part of the Escambia County School District.

Cantonment, Florida, not a part of the original comparison group, was added to the present research to provide a comparison for the Crestview low-income black students. A major research problem in the original Planned Variation research was that in some sites the vast majority of students who were eligible for Head Start were then enrolled in the program. Because of this, comparison groups from a nearby site, often of questionable comparability, were included. Of course in trying to construct the addition a post-hoc, no-preschool comparison group 20 years later, we were faced with the same constraint.

Greeley, Colorado is partly a university town with a large rural area surrounding it. The subjects from the Head Start program included students from low-income Hispanic, largely Mexican-American, migrant families who had settled there and low-income white families. The students went to either a school designated as the Follow Through school or to one of the other schools with children from largely lower-income families.

As indicated above, in addition to the High/Scope and regular Head Start groups, for each Head Start site, i.e., Greeley, Ft. Walton Beach, Crestview, Pensacola, and Cantonment, we developed a no-preschool comparison group. This would allow for a Head Start versus no-preschool comparison and some control for site differences, although we had to use Pensacola to form a no-preschool group for Ft. Walton Beach. The school districts and the Head Start agencies in Ft. Walton Beach and Pensacola, Florida and Greeley, Colorado have cooperated and worked very hard to help us identify and locate the former students we have interviewed.

A total of 625 former students from the two study sites were individually interviewed in person (a few were conducted on the telephone) over a three-year period. The subjects ranged in age from approximately 19 years to 23 years of age at the time of their interviews. At the time of entry, or eligibility for entry, into Head Start, the students' ages, were from four to five years, spanning a 24-month range. This problem is also a potential problem in the Follow Through study under consideration, especially if 2 cohorts (1970-1971 & 1971-1972) are used. Although we anticipate that age at the time of interview will be distributed similarly across groups, we also expect that at least some analyses will need to be conducted with age as a variable. We were faced with many constraints that resulted in the three-year interview phase. A feasible and better time frame would be two years.

Issues and Problems in Forming a Comparison Group

In the proposed Follow Through Study, as in the Head Start study, even if subjects selected for the nontreatment comparison group are well matched, we cannot overcome self-selection and other potentially biasing factors. There may be something different about the family backgrounds, personalities, or motivations of students who enrolled in the program, compared to students who were eligible, but who did

not enroll. In the case of Follow Through, these reasons most likely pertain to where and why the school system located the program.

In trying to locate and involve former students in the Head Start study, we also found that we had better contacts and more help. We presented the study as a follow-up of former students from the school district who may have been in a variety of school and preschool programs. Still, the families with children who had been in Head Start were more willing to provide location information, especially in those cases where we had someone from Head Start locate them. We tried to minimize direct contact with Head start staff, except when we got down to the hard-to-locate subjects. Thus, even in the process of conducting the research, it was hard to avoid self-selection factors.

Potential bias is a common limitation in studies of programs designed to serve as many students as possible. When the primary purpose of programs is mainly to provide a service, not to conduct a research study, random selection of participants is usually not feasible or desirable. Nevertheless, the primary goal in the selection of the comparison group is to form a highly similar group that will provide a valid basis for comparison.

Size of a Follow Through (FT) study sample

One proposal under consideration is that a follow-up FT study include ten major, original sponsors with five sites per sponsor, resulting in 50 sites with a minimum of 100 subjects per site (50 FT subjects and 50 comparison subjects). This would mean 5000 subjects, 2500 FT and 2500 comparison subjects. The proposed criteria under consideration for the FT sites and subjects is:

- Students who attended a full term of the FT program, i.e. three or four years
- Indication of Head Start Planned Variation and Follow Through
- Not more than two projects per sponsor, per region
- Mixture of urban/rural sites per sponsor
- Variability of ethnicity
- Students from the 1970-1971 or 1971-1972 FT cohorts
- Students with kindergarten or first grade entry into FT
- Availability of comparison groups
- Adequate site implementation ratings for a model

It has been anticipated that many or most of the FT sites would include a newly-formed comparison group. Major difficulties will develop if a given sponsor or given site does not have records that are adequate to establish the above factors to be used as criteria for the selection of the comparison group, although my impression is that the sponsors do have good identifying and background information on their participants. If the sponsor has the original lists of names with the eligibility indicated for each family from the program and the original comparison groups, then

there would seem to be an adequate basis for determining the appropriateness of the original comparison group or establishing a new one.

As indicated, there were various problems with the original Planned Variation comparison groups. In some cases, there was potential contamination since comparison subjects were sometimes in neighboring classrooms. There were school differences, with some comparison students in highly similar schools to the FT students and others not. Some comparison schools or sites were different from the FT ones in urban/rural, socioeconomic, or ethnicity factors. Also, Follow Through groups usually included a mixture of students with and without Head Start experience, but preschool experience of the comparison group was not necessarily well matched to the FT group.

Using school records and 1970 census data

If eligibility of potential comparison subjects is not known from program records, school records should be helpful. School records often, but not always, include parent occupation. One limitation is that the record may include the occupation of only one parent, even though the other parent worked part-time. The job titles are also a limitation since often the place of work, rather than the position, may be indicated. Other times the nature of the work, e.g., sales, factory worker, is too ambiguous to establish level of job status. To overcome these problems in the Head Start study, when we tried to locate former students, we asked the parents we contacted to provide information about their employment and educational levels at the time of the student's program participation. However, we found that asking a few quick questions about their occupation or educational level was not sufficient to overcome the tendency to forget, not answer, or embellish. A more thorough interview with a series of questions should help to overcome this, e.g., using a life-span interview approach.

In the Head Start study, the school records did contain addresses of families at the time of the students' school entry. These addresses were used to identify the neighborhoods, which were highly indicative of socio-economic background as shown in the 1970 U. S. Census Bureau data. To select a no-preschool group, the addresses and schools of the Head Start students were used to establish a pool of potential comparison students.

In Pensacola, we first identified the census tracts (blocked by the 1970 U. S. Census Bureau data) in which the Head Start subjects lived in 1971-1972 (year of entry in public school). Using the school records data, we established which public schools the Head Start students had entered. Using the school-entry addresses of both Head Start and nonHead Start subjects, we identified all no-preschool, nonHead Start students who also lived in the same tracts or areas as the Head Start students.

To achieve similar Head Start and comparison groups with matched proportions on major demographic variables, students were blocked according to school, gender, ethnicity, and same census tract or street. In Pensacola, since there were more eligible subjects than needed, students were then randomly selected from each appropriate block to fill the needed proportions. In Crestview, Cantonment, and Greeley, which had not been tracked by neighborhood by the Census Bureau, we used the criterion of same street, closest address or neighboring street.

Size of comparison samples

There has been the suggestion that size of samples in a national FT study would be fixed. If conclusions are to be made at the site level, it may be more desirable to select a sample size according to how representative a study sample size would be of its original group. For example, one site may have a larger original sample in a given region, e.g., 150, where n=50 would represent only 33 percent of the original group, whereas another original group of only 75 would have a much more representative group with n=50 (67 percent). However, weighting of data from samples may be appropriate, but these matters would need to be worked out with sampling experts (e.g., Kish, 1965).

In the Head Start study, since we wanted to be in a position to make conclusions at the site level, we strived to locate and interview the majority of the original Head Start groups and the newly-formed no-preschool comparison groups. A comparable number of no-preschool subjects was also desirable at each site. We therefore targeted the comparable population sample and over enrolled subjects for the comparison group. Thus, subjects with inappropriate socio-economic family background could be later eliminated on the basis of information from the interview. The interview data is also a basis for eliminating comparison subjects who indicate that they had attended some other kind of preschool program.

In general, in order to achieve representativeness of the program and comparison groups, all individuals in the sample pool needed to be actively pursued. It is important to interview a roughly equal proportion of "hard-to-get" students in program and comparison groups, even though we may over enroll the comparison group to ensure a sufficient number or to maintain some extra subjects for final matching of the groups.

What is a final acceptable number of subjects for the FT comparison groups? The goal is to achieve a representative, comparable number of comparison subjects, nearly equal or at least proportional in cell sizes, as has been suggested for the FT study. For example, if 50 program subjects is determined to be the desired cell size range to be achieved, we can expect that 60 to 75 percent of any pool of potential comparison subjects may be found and subsequently actually are interviewed. We know that the pool of potential subjects for the comparison groups may need to be larger in case they are harder to locate or involve than the former program participants. These feasibility considerations indicate that 100 comparison subjects would need to be searched for in order to ensure that 50 to 75 subjects would be available for the final study comparison group.

Location and Interview Response Rate Problems and Strategies

Locating, contacting, and finally actually interviewing subjects in the Head Start study was more difficult and complicated than anticipated. Even where the schools supplied one- and two-year-old addresses and telephone numbers for the graduates, many of these addresses proved to be out-of-date with no leads to follow. School personnel were not in a position to devote the time and effort that was needed to trace former students through family, friends, and associates.

BEST COPY AVAILABLE

The majority of students were found to be still living near their original address, in the next town, or in a different part of town. Others were away at college, in the military, or married to someone in the military, either in state or out of state. Others were in the same general region of the state, but in another city. The hardest ones to find were those who were not in close contact with the family or were living with someone on a temporary basis. Although some students had moved away with their family during their school years, there was often someone who knew where they moved to.

Very often when we would get a good address for someone, they would often move or their telephone would be disconnected as soon as several days later. In part, such quick changes in addresses and other location problems are due to the mobility of the age group of these young adults. Other problems, often cited in other research studies conducted in low-income communities, were that telephones are frequently disconnected because bills could not be paid, people are evicted or have chosen to move for economic reasons, younger adults move in with others when they run out of money, and people may be avoiding bill collectors, welfare department case workers, the police, or personnel from the school system. Due in part to such factors, there is often a level of mistrust in low-income communities that has to be overcome. We used many strategies to deal with the mistrust, but not one of them worked with everyone. Nevertheless, we also found that most of the people we contacted, especially the parents, were very interested in the project and very responsive.

We selected a diversity of locators and interviewers many of whom were from the same general ethnicity of a given community. All the interviewers had college-level education or considerable professional experience working with people. Interviewers had to be able to master the content and format interview questions, but still interact in a natural, but professional manner.

Specific strategies for location of subjects and scheduling interviews

Local Project Coordinators. The basic system we developed in the Head Start study included hiring a part-time local study coordinator in each site working under the direction of the High/Scope project director who supervised them by making site visits, frequent telephone calls, and through mailing specific instructional procedures. For a FT study, I would recommend that full-time coordinators be hired to maximize the effective management of the coordinators' time. The first job of the coordinator was to assist in the hiring and training of locators and interviewers from the community. The coordinator also provided the direct supervision of those working on the project.

To minimize bias, coordinators and interviewers were not present or past employees of Head Start. The coordinator should be either someone with research experience or someone with considerable contacts in the community, specifically involving one or more community networks, i.e., the school system, the Human Resources Department, etc.. The coordinator has to identify key community leaders and long-term neighborhood residents who can provide indirect and in some cases direct information, addresses, and telephone numbers of families and young adults who used to live in the neighborhood.

Although it may be preferable to hire a coordinator who presently lives in the key neighborhoods, we also found coordinators who were effective who did not live in the neighborhood, but were experienced in working with a diversity of people. Coordinators, locators, and interviewers varied in ethnic backgrounds, but often were not of the same ethnic background as the families they were trying to locate. Locators were often also interviewers, but some locators were not judged to be effective, appropriate, or were not interested in being interviewers and thus only conducted location work.

To protect confidentiality and minimize potential bias, the people hired to be interviewers did not know the people they were interviewing or their families. For similar reasons, coordinators usually did few interviews since it was hard to keep them unaware of which subjects were in the Head Start and comparison groups, which interviewers typically did not know. The coordinator would also often get to know more about the subject's background and even sometimes get to know the subject somewhat in trying to schedule the interview.

Establishing a Resource Network. I recommend that a Follow Through Study employ a strategy of establishing a community network. To establish a community resource network in the Head Start study for the location of subjects, former program directors, teachers, teacher aides, and other staff (e.g., parent coordinators); parents, formers students, and key community leaders from 1970-1971 were identified and located, with assistance from current staff from Head Start and the school districts.

Former Program Directors, Teachers, and Coordinators.The former key staff members from the Head Start program and FT in the schools were located where possible. To find some former teachers who had moved, we obtained information from the school district personnel files on the college they had graduated from and we then contacted the alumni offices to locate them. The project director or the coordinator met with them or talked with them by telephone to gather lists, notes, and photographs of former students from the High/Scope model and regular Head Start classes, and the comparison group.

In the Head Start study, some of the current or previous program staff members agreed to visit the neighborhoods to check on addresses and telephone numbers or the general current whereabouts of family and former students. Thus, current and former staff people, along with other key identified community leaders, would help to provide a resource network for the coordinator and locators to call upon for clarification and leads for locating and often relocating former students and their families.

For the FT study, the former and current FT teachers could also be interviewed formally to provide a more full, historical qualitative description of the program, including a focus on the perceived immediate and subsequent impact of FT on the teachers and others in the school district, e.g., stimulating other educational programs and methods.

Student and Parent Directory Information. Directory Information from Schools. In general, directory information is available from program and school records, although schools varied in who they allowed to record the directory information. Many schools require their own staff to gather the information. Some school districts did have computer printouts with directory information from the time of the

118 BEST COPY AVAILABLE

students' entry and their last year of enrollment. Key identifying information should be carried forward on lists used by the research staff, specifically:

- the full, proper name of the student ,
- any name changes, due to adoption or marriage (new names should always be included with the original, using a consistent order),
- student's birth date,
- student's gender,
- student's ethnicity,
- name of school first entered,
- name of school graduated from, last attended and/or transferred to,
- parents' addresses and telephone numbers—both at the time of school entry and last recorded

If FT programs have most of the above information from the FT student program lists, the information can be verified and addresses and telephone numbers added. In the Head Start study, once we established the lists of Head Start and Head Start/FT students, we compiled lists of students and addresses of all students who had also entered the same elementary schools at the same time that the Head Start students entered in 1971-1972, or entered schools identified as having students from comparable socio-economic and ethnic backgrounds. As described previously, the school, gender, ethnicity, and neighborhood were then used to select a comparison group in matched proportions to the Head Start group in each site. The lists of school-entry and last-recorded addresses were then also critical for location purposes. Yearbooks and class photographs were also available from the schools and were sometimes useful for student identification. Also, yearbooks are a good source of further information on student activities.

Public Records. There are several accessible public records sources that were used in the Head Start study to find parents and former students. Marriage, birth (for children of former students) and voter registration records were each used, after other resources had been exhausted. The most valuable resource, other than the school, was the State Bureau of Driver's Licenses, which was accessible at the state level for a small fee for each individual person in the computer search. In some cases, the fee was waived. The full name and birth date of a subject was required. Some searches also require the social security number, which we did not have. Parents and/or students can be searched for, and the most recent address and often the telephone number are thus obtained. We also used this resource when we knew only that a family or student had moved to a certain state.

If social security numbers were available to FT sponsors, I understand that credit searches could be conducted, although many former students and many of their parents will not have a credit record. I also have been told that some federal offices, e.g., the Social Security Office have sometimes agreed to forward mail to someone if a social security number is provided.

Institutional Searches. To determine if some of our former Head Start study subjects were in prison, local coordinators worked with state and federal prison administrative staff who used their computer systems to search for names from our lists. Local

coordinators then arranged to contact those former students found to be in prison and a number of interviews were conducted in prison or after the subject left prison. Military locator assistance on military bases was also used in states where we learned that subjects were stationed. For a national FT study, major central searches in both the prison and military systems may be feasible. Coordinators in the Head Start study also contacted mental institutions in a few cases where it was learned that one of our students might be located. Some directory assistance was provided from some human resource departments who had current addresses of study subjects who had applied for assistance in finding employment or were enrolled in a publicly-funded work program. Assistance from public financial assistance programs, e.g., AFDC was not made available to us.

Local coordinators also checked with major employers, e.g., factories, businesses, and local colleges to locate study subjects or parents. Churches in the major neighborhoods were sometimes able to provide a recent address of a church member. When we knew someone was at a certain out-of-town college, we contacted the college to get an updated telephone number.

Public Announcements. Announcements of the Head Start study were posted (as school follow-up studies) in the churches, schools, Head Starts, and community bulletin boards with the coordinator's telephone number to call. We made general announcements of the study on the radio and in various local newspapers. In Greeley, the study was also announced on the radio in Spanish. These strategies were helpful in some few cases, but many people would call who were not eligible for the study. However, these strategies did help to create a community awareness, legitimation, and acceptance of the study and the interviewers. Whenever locators and interviewers were in neighborhoods for the first times, the local police department was notified so that the study could be verified by anyone with a question. Similarly, contact persons at Head Start and the school systems were identified for interviewer verification purposes. All interviewers carried official High/Scope identification.

Telephone Directories. In addition to the standard telephone directories, several kinds of other directories are useful. The most useful directories are those also used by businesses and are obtainable by contacting the directory companies or a city's chamber of commerce. These directories are cross referenced by street as well as by name. Thus, streets can be reviewed for resident names of former students, family members, or relatives. In some directories, a person's place of work and work telephone number are also provided. When a telephone number is disconnected or no longer had the same party, local and regional telephone operators are also helpful in providing recent telephone numbers. Many times coordinators and locators would finally locate someone by calling everyone with the same last name in an area and thus find a subject's relative.

Mailings. Mailing information about the Head Start study was most often used to provide an advance introduction prior to the interviewer's calling to schedule an interview, provide verification for uncertain parents or students, to confirm a scheduled interview, or to indicate regret if a subject did not show and to request for another time. Telephone calls were also used for confirmation of interviews and checking on rescheduling if a subject did not show or canceled.

Several kinds of letters were used to explain the study to potential participants. For students where we had an address, but not a telephone number, we were

instructed by the U. S. Postal Service to write "Forwarding Address Requested." When there has been an address change, the new address will be recorded and the letter returned to us. We then used this new location information to pursue the subjects. Although most of our letters were brief, simple, and readable, for people with more questions, we did use some letters that explained more of High/Scope's background and experience and provided some more information on the research. In these letters, however, we were careful not to create some context that might frame future types of responses of interviewees.

Parents who refused to give out their son or daughter's address or telephone number would often agree to forward a letter that we sent if we also provided a stamped, enclosed envelope for this purpose. Some parents would also agree to ask their child to call us. These second-party strategies resulted in information getting to the student about half the time. We also enclosed postcards or a form with the letter to parents and students with spaces for telephone numbers and best times to call. We also had a postcard for out-of-town subjects to send to us prior to coming home for a visit so that we could arrange an interview. This letter strategy only worked occasionally. However, many interviews were successfully prearranged by parents to coincide with a son's or daughter's visit from out of town.

Neighborhood Canvassing. Coordinators, locators, and interviewers, canvassed key neighborhoods, gave information about the study, and showed residents the lists of former school district students not yet found. Sometimes a Head Start or a school staff employee who had a well-known face in the community would accompany the coordinator to provide introductions and assurance. This step was especially employed for the final hard-to-find subjects. Since many low-income families do not have a telephone, some neighborhood or specific home visits are necessary.

Interview and Contact Procedures and Obstacles

Many approaches were tried. The basic approach was to have some people assigned to do location work under the supervision of the local coordinator who also did some location work. Location work consisted of building and tapping into contacts in community networks, inquiring about former students and their parents, and verifying addresses and telephone numbers. This information, including more vague leads was given to coordinators. The coordinators found the first available interviewer, preferably of the same gender and ethnicity, who would then contact the potential subject, explain the study, answer questions, and if possible schedule the interview. This process worked for the majority of the subjects, *but not for those who were to difficult, reach, to schedule, or to actually be interviewed.*

The major problem was that many potential subjects agreed to be interviewed, but would not show up several times, yet would not just refuse. Most subjects were interviewed as scheduled, but a sizable number had to be rescheduled several times before it actually worked out without their canceling or not showing up. Some subjects would be interviewed only after a period of one to several months had passed. Coordinators and interviewers would periodically recontact some "hard-to-schedule" subjects and try to again schedule the interview. Often simply catching a subject at a less busy time, a subject would then be scheduled and finally interviewed. Another approach was to have interviewers who were quickly ready to go do an interview. When contacting a subjects was successful, the coordinator could then set

up an interview to be done that day of the next before the person got busy or changed his or her mind. About five to ten percent of those former students whom we directly contacted did definitely decline to participate, after several periodic appeals for their participation.

Student and parent contacts

After students were interviewed, particularly students who seemed to know many other students, the coordinator and sometimes the interviewer, would show a list of not-yet-located subjects to the interviewee to inquire whether they knew the whereabouts of the student or a family member. If an interviewee's mother had been active in the school or community, she would often also be a good resource for the coordinator to check with to get leads on the location of other students. Inquiring of subjects and parents for leads to use to locate other former students was a fairly helpful strategy.

Out-of-Town Subjects. Approximately eight to ten percent of the overall sample lived away from their original town or city, and another 15 percent lived out of their original state. We arranged several trips to other parts of Colorado and Florida and nearby states in the South. We also set up interviewing in Texas, North Carolina, California, and Arizona, and interviewed several subjects in military bases in Germany. A small number of subjects were interviewed on the telephone. The coordinators made the arrangements for the out-of-town trips and often accompanied the interviewers or in some cases actually had to do the interviewing. For security reasons, the coordinator or interviewer always had another adult, either another study staff person or a spouse, accompany them on out-of-town interviews.

Interviewee Compensation. After being interviewed, the interviewees received a $25 check in the mail for compensation for their time for completing the interviews, and this incentive appeared to be positively received by both Head Start and comparison subjects.

Target location and interview response rates

The overall rate of located subjects exceeded 80 percent and in some sites 90 percent. The overall interview response rate, i.e. those subjects who were actually interviewed ranged from approximately 73 to 90 percent across sites and subgroups, e.g., male, white, Crestview, with an overall average of 80 percent.

I agree with the report by Lally, Mangione, and Honig (1988) that the last 10 percent of the subjects we interviewed were living in unusual or difficult circumstances, i.e., avoiding bill collectors, involved in selling drugs, avoiding an abusive spouse, etc.. These circumstances were also found among earlier-interviewed subjects, but less often or less intensely.

There was nothing obviously different about subjects in the Head Start study who were interviewed in the 70 to 80 percent range compared to the overall sample, but we have not analyzed the data yet to examine this. It was our general impression, however, that the first 50 to 60 percent of those located were in relatively more positive circumstances. Nevertheless, it was very hard to actually locate and

contact families or former students who had simply moved away, and these subjects were often in very positive life circumstances.

It should be noted that subjects in prison or recently in prison were not difficult to locate due to the directory assistance of the state and local systems. However, we did not survey states outside of Florida and Colorado unless we had a lead that a subject was in prison in a specific state. In a national FT study, tapping into the national network computer system would be advisable.

In contrast, when someone had a friend, relative, or community person who knew that former student was in college, they were usually eager to tell us where they thought the student was. I do not know what comparable networks there might be for students enrolled in colleges, but we only pursued those for whom we had a lead.

Interviewer training

Interviewers were trained by the project director using an interviewer manual, demonstrations, and role playing. Interviewers were required to conduct several pilot interviews and to tape record them. These audio tapes were reviewed by the project director who then provided feedback. Some interviewers were not hired after it was determined that they would be too difficult to train. Interviewers were largely hired upon recommendations from professionals in the community. Many interviewers did only a small number of interviews because they found it too difficult to schedule an interview or had too many interviewees not show up. The completed interviews were turned into the local coordinator who checked only the clarity of handwriting and sentences.

The interview instrument

An interview instrument in the Head Start Study was set up in questionnaire format and was the major methodology used to gather the data. As indicated, the interviews were conducted in person. With individuals who may have difficulty in comprehension or establishing rapport, the interviews are typically conducted in person. The content areas for the interview questions regarding the subjects included:

1. family background from childhood to the present, including education, employment, family composition, type of household, and perceptions of parents' attitudes;
2. current family status, including marriage, children, living at home or elsewhere;
3. experiences and achievement in school, work, child-rearing, and the community including contributions to the community, citizenship, and delinquency or criminal activity;
4. economic sources of support including self, family, government, other;
5. attitudes toward self, work, education, children;

6. perceived sources of influence, including family, school, and other experiences; and
7. future goals and plans.

Over one half of the questions reflect content from 1, 2, 3, and 4 above. These questions were designed using methodology previously developed and employed with adolescents and young adults over the last ten years in ongoing long-term studies where the long-term effectiveness of High/Scope's educational programs have been examined. The other related studies include a follow-up assessment of High/Scope's former Perry Preschool children, at ages 15, 19, and 28 years, and the long-term follow-up assessment of former Head Start children as young adults. Over the years, the content and focus of the items, formats, and survey methods have been designed based on: reviews of the relevant theoretical and empirical research; findings obtained from ongoing research at High/Scope, e.g., Berreuta-Clement, Schweinhart, Barnett, Epstein, and Weikart (1984); consultation with researchers from University of Michigan's Institute for Social Research (ISR); and from pilot work we have conducted specifically for each study.

To evaluate subjects' comprehension of the interview items for the Head Start study, the interview instrument was piloted with approximately 50 pilot subjects and subsequently refined. The final instrument, with newly developed items and revised items, was then piloted with additional pilot subjects. Revision of the content and format of the items were then finalized.

The major research questions of interest in the Head Start study include whether or not the Head Start program, and the High/Scope model in particular, influenced subjects' later situations and behaviors, particularly in the areas of educational level, employment, or economic self sufficiency.

Records Data Collection

Records data are also important to collect, to provide some external, independent data sources in addition to the self-report data gathered in interviews. Such information is also useful for a cost-benefit analysis as conducted in the Perry Preschool study. In that study, analyses of the data indicated that attending preschool was more cost-effective in the long run in terms of less actual costs in criminal activity and economic independence. For the present Head Start long-term study, this information is being collected within a fairly narrow time frame so that subjects could be compared within the same time frame.

Educational records

In the Head Start long-term study, we are just completing the collection of educational records data, including all school records from kindergarten through high school, and for the majority of subjects, any postsecondary records, including GED, adult education, college and technical courses and degrees up through to the present, early 1991. School districts, including out-of-town districts, have provided the information. The Department of Education in Florida has coordinated a state-

BEST COPY AVAILABLE

wide search of student records to achieve the postsecondary assessment. We will also pursue out-of-state education, but this is not finalized. Signed informed consent was required to access school records information.

Criminal records

Basic criminal records information is generally open to the public, but cooperation is necessary with local and state court personnel to get information from records or computer systems. We have not been able to secure computer printouts that were complete at the state level without prohibitive cost as yet, but this may still work out. For a national FT level study, the federal computer system should be the major information sources, although some levels of criminal activity will not generally be forwarded to that level. The local records are sufficient for criminal activity in the county or region, but this does not include crimes committed elsewhere.

Public assistance records

Data collected from records of AFDC, food stamps, and other sources of public assistance data collection are still underway in both sites, but will soon be completed. Signed informed consent is required to access these records data.

Summary of Key Issues Regarding Feasibility of FT Follow-up Research

There are several major methodological issues to be examined in determining the feasibility of a national follow-up, Follow Through Study:

1. *Selection of Sites.* Evidence of adequate FT program model implementation and sufficient lists of students and other background data on students should be two major considerations in model and site selections.

2. *Selecting Follow Through and Comparison Groups.* The criteria should take into account the representativeness of the groups at the program and site level if generalizations are to be made at that level. Sampling experts should be consulted for technical assistance. An alternative would be a national random sample generated to make generalizations of Follow Through without regard to program model site effects.

3. *Finding Former Students.* There are location challenges in trying to find subjects who have not been tracked as in planned longitudinal research. However, even in low-income neighborhoods, it is feasible, using well-known strategies and as well as strategies High/Scope implemented, e.g., setting up community resource networks.

4. *Interviewing Subjects.* There are very difficult challenges in gaining subjects' interest, trust, and following through with scheduled interviews. Although subject alternates are possible for the comparison group, the response rate is really based on the overall pool to be searched for. Young adults in this transitional age group and lower-economic families are very mobile, have busy, often unstable circumstances, and/or are trying to avoid people they

do not know who may be from the welfare department, police, or bill collectors. Although the Head Start study was conducted over a three-year period, a two-year period is feasible. Strategies have been identified to reach a highly representativeness interview response rate. However, a FT national, multi-site study would be expensive if done with a large sample over many sites. Interviewing parents and former/present FT staff is very feasible given that, compared to younger adults, they are easier to locate and tend to be less mobile once they are located.

5. *Records Data.* The availability and collection of records data, i.e., school, criminal, and public assistance records is challenging because some information in files has been deleted, completely destroyed, or has been summarized to a very minimal level. Some data might be available from some FT program sponsors, which would cut down on the expenses and limitations of totally new records data collection on site.

6. *Costs for a Multi-site Study.* For the Head Start study, High/Scope was able to conduct the subject location and interview data collection at two major sites for what amounts to approximately $500 per subject. This includes a considerable amount of donated time by various administrators and staff at High/Scope and the sites. For a multi-site FT study, there should also be a coordinating center to ensure common procedures across many sites. It is estimated that a multi-site study larger than High/Scope's study would be more costly to organize and implement. Costs for records data collection also need serious consideration.

Estimated costs for a multi-site Follow Through study

The cost estimates for a multi-site, follow-up FT study are based on developing, conducting, and analyzing the data for approximately 600 final study subjects. In general, we estimate the following per-subject costs:

Location and interviewing phase:	$800 per subject
(Location and interview phase would include some staff at the Coordinating Center, employment and travel of the Project Director, all Local Study Coordinators, locators, interviewers, and interviewees and institutional overhead.)	
Coding, entering, and analyzing interview data:	$200 per subject
Records data collection:	$250 per subject
Coding, entering, and analyzing of records data:	$200 per subject
TOTAL COSTS =	$1,450 per subject

Given the projected 5,000 FT subjects for the national Follow Through Follow-up Study, it is thus estimated that the study would cost $7.25 million. If some percentage of parents and teachers are interviewed, for example, 2,500, the costs for interviewing, coding, entering and analyzing the data would be approximately $1,000 per subject or $2.5 million, in total, although this is harder to estimate. The

BEST COPY AVAILABLE

study would appear to cost in the $8 to $10 million dollars range, although it may be more expensive depending on the main contractor and any subcontractors that would be necessary to launch the study at a national level.

References

Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives. The effects of the Perry Preschool Program on youths through age 19. Monographs of the High/Scope Educational Research Foundation*, No. 8. Ypsilanti, MI: High/Scope Press.

Guenzel, P. J., Berckmans, T. F., & Cannell, C. F. (1983). *Economic survey methods*. Ann Arbor: ISR Publications, Institute for Social Research, University of Michigan.

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.

Lally, R. J., Mangione, P. L., & Honig, A. S. (1988). The Syracuse University Family Development Research Program: Long-range impact of an early intervention with low-income children and their families. In D. R. Powell (Ed.). *Parent education as early childhood intervention: Emerging directions in theory, research and practice. Annual advances in applied developmental psychology*, 3. Norwood, NJ: Ablex, 79-104.

Lansing, J. B., & Morgan, J. N. (1971). *Economic survey methods*. Ann Arbor: University of Michigan Press.

Lazar, I., & Darlington, R. (1982). *Lasting effects of early education: A report from the Consortium for Longitudinal Studies. Monographs of the Society for Research in Child Development*, 47 (2, Serial No.195).

Ross, R. T., Begab, M. J., Dondis, E. H., Giampiccolo, J. S., Jr., & Meyers, C. E. (1985). *Lives of the mentally retarded. A forty-year follow-up study*. Stanford, CA: Stanford University Press.

Seitz, V., Rosenbaum, L. K., & Apfel, N. H. (1985). Effects of family support intervention: A ten-year follow-up. *Child Development*, 56, 376-391.

Smith, M. S. (1975). Evaluation findings in Head Start Planned Variation. In A. M. Rivlin & P. M. Timpane (Eds.). *Planned variation in education*. Washington, DC: Brookings Institute, 101-111.

Recommendations

Robert Egbert
University of Nebraska

Edgar Epps
University of Chicago

Garry McDaniels
Skills Bank Corp.

In order to help provide information essential to educational policy makers, planners, and local schools as they strive to reach the goals outlined by the President's Education Summit, we recommend that the Department of Education conduct a long-term Review of the Follow Through Program. Specifically, we recommend that the Department of Education take a leadership role, that would also involve the Department of Health and Human Services and the Department of Labor in planning and conducting a comprehensive, coherent set of follow-up studies of Follow Through. These studies could provide a basis for fundamentally altering the organization of services for low income three to nine year-olds in the United States.

The timing of this proposed review is critical for two reasons: (1) the information is needed now, and (2) the data sources needed for such a review are available now, but they are disappearing. We know for example that graduates of the Follow Through program—as well as their parents, teachers, and counselors—can be located (Oden, 1991) but the task becomes more difficult as time passes and the graduates get older and more dispersed and as teachers, counselors, and other informants and participants retire or otherwise leave their jobs. We also know that a great deal of information is available with program sponsors as well as with local projects, but records are gradually being destroyed and discarded, computer data

tapes are aging and those tapes are becoming less readily interpretable. In short, now is the time to conduct a Follow Through Review.

Background

Project Head Start, funded under the Economic Opportunity Act, was introduced in the summer of 1965 as an experimental, summer program of comprehensive services to children. When end-of-kindergarten follow-up studies of first summer Head Start graduates suggested that some of the achievement test gains that the children made during the first summer were subsequently lost during kindergarten (Wolf & Stein, 1966), President Johnson requested a program that would preserve and extend the gains of Head Start. Like Head Start, this Follow Through program was to provide comprehensive services; it was to do so for early elementary school children who had graduated from a Head Start or similar preschool program.

The federal budget revisions of FY 1968 brought major reductions in a number of Great Society programs including a decrease in the Follow Through budget from $120 million to $15 million. At that time, the conception of Follow Through was changed from that of a large, operational program to a developmental or experimental program that would enable learning more about the outcomes which comprehensive services combined with different curricula and teaching strategies might have with young, poor children. A program and research approach termed planned variation, in which local projects were associated with institutions that had developed promising approaches for the education of young children, was prepared for Follow Through and put into place in the fall of 1968 (Egbert, April 1973).

Comprehensive services within Follow Through included instruction; medical, dental, nutrition, social, and psychological services to children; and parent education. In addition to parent education, Follow Through sought to have parents participate through service as advisory committee members as well as being classroom, health, social service, and parent aides. Parents were encouraged, through home visits and the provision of parent rooms as well as by frequent socials and coffees, to trust their schools as well as to be present in them. Parents were encouraged to take more control of their own lives as well as the education and care of their children.

Like Head Start, Follow Through was authorized under the Economic Opportunity Act, but program operation was delegated from the Office of Economic Opportunity (OEO) to the Office of Education within the Department of Health Education and Welfare (DHEW). Follow Through funding grew rapidly during the first three years. Program funding leveled off during the initial research and evaluation period and then decreased. Although Follow Through has continued as a program, its funding has decreased, in constant dollars, to less than ten percent of what it once was.

Numerous federal, state, and both public and private local agencies participated in Follow Through. Among others, this included state education and economic opportunity offices, local school districts, private schools, medical and dental clinics, universities, community and health agencies, and regional educational laboratories.

Graduates of the Follow Through program are now young adults, and it is possible to study indicators of the quality of life that they are experiencing. Among these indicators are the education that they pursued and the success that they experienced

BEST COPY AVAILABLE

in their education; the frequency of early school leaving, delinquency, and arrests; and their employment and health status.

Potential Value of the Proposed Follow Through Review

Political, educational, and business leaders at national, state, and local levels are seeking ways to improve the education of America's children and youth. Diverse efforts range from revisions of school curricula, to changes in the decision making structure within and among schools, to changes in instructional methodology. These reform efforts are both well intentioned and based on rational interpretations of child development and education; nevertheless, they all suffer from the gaps that currently exist in our knowledge about the long lasting effects of various early elementary and preschool programs—effects on the children who were enrolled in those programs and on the adults (teachers, aides, parents, and others) who worked with them as well as on the schools and other institutions. Ultimately, the efforts of research about the Follow Through program and its graduates could result in improved educational programs for young students. This is not to suggest that this research would, or should, produce a single formula. Rather, the intent of the research would be to help us be as informed as possible about the variety of long term outcomes of early elementary education.

Information gained from a Follow Through review could be especially useful to AMERICA 2000 on at least six dimensions.

1. *Choice.* As perhaps no other educational experience, Follow Through offered the parents and school-site professionals the opportunity to make informed program choices. A thorough, prompt follow-up review of Follow Through could provide extensive information about program variation possible at the elementary school site level. This information should assist parents in making informed choices for their children.

2. *The First 535+ New American Schools.* Many of the first 535+ New American Schools should be elementary schools. Those who plan these schools would benefit substantially from information produced in a major Follow Through review. Such a review could provide information about program outcomes, parent involvement, teacher development, site-based leadership, and other crucial aspects of program planning and evaluation.

3. *The School As the Site of Reform.* Follow Through projects provided both education and comprehensive services under local, site-based leadership with parent involvement. Follow-up studies of how this influenced project service and human outcomes can inform those working with the schools as the site of reform.

4. *Promote Parents' Role in Their Children's Learning.* Parent involvement, including involvement in their children's learning, has been one of Follow Through's most prominent features. Some sponsors have been particularly strong in this emphasis. A Follow Through review could both identify effective procedures and provide information on the overall effectiveness of these attempts.

5. *American Achievement Tests.* Third grade achievement test information is available on thousands of Follow Through graduates. Examination of these scores against later school and non-school performance could give useful guidance to those responsible for developing the American Achievement Tests.

6. *Prepare Report Cards That Report Student Performance.* Follow Through program sponsors have had extensive experience in developing novel and useful reporting formats. Because program emphases differ depending on the program approach, having varied reporting formats also is important.

The Follow Through Program provides an ideal opportunity to study the long term effects of various approaches to educating young children. The Follow Through sponsors, projects, and children present the largest sets of projects, schools, and children on which data are available. In some instances, a great deal of program process information also is available on these projects.

A well-designed, coherent set of studies about Follow Through could enable policy makers, planners, and local education leaders to improve dramatically the quality and organization of services for low income, three to nine year-olds in the United States. Positive data could argue for

- increased coordination of educational, health, and social services.
- greater local, including site level, leadership in choice of education approach and control of services.
- increased emphasis on employment of parents in the education, health, and social services process.
- better coordination of, or a single local manager of, Head Start and early grade programs.
- improved state coordination of higher education resources and early childhood/early grade service providers.

Follow Through Program Review Opportunities

A single major evaluation of the Follow Through program was reported in 1977 (Stebbins, et al.). Several critiques of the evaluation and critiques of the critiques appeared over the next few years (e.g., Haney, 1977; House, et al., 1978; Kennedy, 1978; Bereiter & Kurland, 1981-82). A number of smaller scale evaluations also were completed (Egbert, 1991). Through the years, in addition to the national evaluation efforts, several follow-up studies have been conducted by program sponsors and others (Becker & Gersten, 1982; Clark, 1975; Gersten, Carnine, & Keating, 1984; Cloud, Rentfrow, & Hildebrandt, 1980; Olmstead, 1987). These studies generally have found greater school persistence and fewer dropouts among Follow Through graduates than among comparison groups. Some integrative analyses also have been reported (Rhine, 1981; Wang & Walberg, undated; Wang & Ramp, 1987). No systematic follow-up work has yet been completed however.

BEST COPY AVAILABLE

Present knowledge about long term outcomes of early elementary education is as limited as our knowledge would be about long term social and behavioral as well as academic outcomes of the preschool education of disadvantaged children if we did not have the information from the Perry Preschool Project (Berreuta-Clement, et al. 1984). The results from the Perry Project consistently show that many critical outcomes of early intervention programs cannot be identified until the students are much older. In fact, researchers and preschool educators are continuing to gain more and more useful information as each succeeding set of data on Perry graduates is gathered and reported. The proposed Review of Follow Through would address many of the issues for elementary educators that the Perry Preschool Project addresses for preschool educators. (Congressman Kildee raised this issue during a hearing on the reauthorization of Follow Through.)

Students who entered Follow Through in 1971 are approximately 25 years of age. It seems appropriate that we conduct follow-up studies with them as well as with parents, teachers, schools, projects, and program sponsors to investigate what effects, if any, the program had on children, families, and other human and institutional participants. Conducting the proposed review will be very complex. Only through thoughtful communication between researchers, interviewers, program sponsors, parents, students, school officials, and government agencies can the effort succeed. However, the results should dramatically increase our knowledge of elementary education and provide the basis for improving educational opportunities for all children.

Issues That Should Be Considered in the Follow Through Review

Based on the papers presented and the ensuing discussions at the OERI February 21-22, 1991, Follow Through Review Conference, we recommend that:

- Immediate and comprehensive steps be taken to identify, assemble, and preserve the Follow Through data base that now remains with sponsors, local projects, evaluators, and the Department of Education.
- An extensive long term follow-up review—made up of a number of smaller, focused studies rather than a single, large-scale study be conducted of Follow Through. The review should feature studies that use a variety of quantitative and qualitative approaches. The studies should be designed to answer such questions as the following.

1. How are the Follow Through students doing as young adults?
   a. Frequency of placement in expensive remediation/special education programs
   b. Frequency of high school graduation and GED completion
   c. Frequency of post-secondary education
   d. Employed/not employed
   e. Self-supporting/not self-supporting
   f. Juvenile and adult delinquency, e.g., arrested/not arrested
   g. Healthy/not healthy

Providing information about these questions will involve several studies of Follow Through graduates. Some of the studies might include comparison groups; others might compare Follow Through graduates with national trends; still others should be case studies in which intervening experiences are examined in some detail. Some of the studies should focus on major cities; others might consider rural areas. Consideration should be given to special populations, e.g., Native Americans on reservations.

2. What remains of the principle Follow Through components?
   a. School coordinated comprehensive services
   b. Community-school integration
   c. Professional training/inservice education

Community case studies could be a primary source of information for these inquires. Those who served as program sponsors might be especially qualified to provide leadership for these studies.

3. What were long term impacts on teachers from Follow Through's teacher renewal strategy?
   a. Continued service
   b. Advancement in profession
   c. Nature of classroom organization and instructional processes
   d. Beliefs about learning and teaching

4. What were some of the long term impacts on parents of Follow Through's parent involvement, parent education, and parent empowerment efforts?
   a. Vocation/socioeconomic status
   b. Community leadership roles
   c. Parenting/grandparenting practices
   d. Support for schools and other social institutions

5. What have we learned from other research about young children that will help guide the Follow Through Review?
   Places to begin:
   a. Werner and Smith, *Vulnerable But Invincible*
   b. Lazar, et al., *The Persistence of Pre-school Effects:..*
   c. Walberg, Syntheses of research on teaching, in *Handbook of Research on Teaching*

6. Why were the within-sponsor variations so great? What happens if the school is considered the unit of intervention?

7. In what ways and why has Follow Through influenced other Federal programs, e.g., Chapter 1, National Diffusion Network?

Suggested Next Steps

A coherent, comprehensive set of follow-up studies of the nature suggested requires that the Department of Education assume the lead role in:

1. establishing initial priorities and creating a tentative list of study areas;

BEST COPY AVAILABLE

2. securing cooperation from Department of Education programs that would both contribute to the research and participate in needed program changes if findings support the need for changes; and
3. securing cooperation from Department of Health and Human Services and Department of Labor programs that would both contribute to the research and participate in needed program changes if findings support the need for changes.

The Office of Educational Research and Improvement recommends that the Department of Education proceed with a Follow Through Review. The first phase of such a Review should involve activities such as:

1. establishing an initial research plan;
2. meeting with affected Department of Education assistant secretaries and the Assistant Secretary for Planning and Evaluation at the Department of Health and Human Services and the Assistant Secretary for Planning at the Department of Labor to describe the opportunity provided by a Follow Through Review as well as for the purpose of getting ideas and setting priorities that would encourage and facilitate participation;
3. commissioning papers and studies that would provide information needed for proposing next steps, e.g., consolidating what is already known about research with early elementary children, and learning what sorts of data are available at Follow Through, sponsor, and site levels (The paper written by Oden is an example of the papers that might be commissioned. Others that should be included are a comprehensive review of studies of effective practices with young, poor children such as the one suggested by Walberg in the February 21 meeting.); and
4. preparing a report that contains recommendations for the remainder of the review. This report should be prepared by September 1, 1992.

Current political and educational reform efforts demand more comprehensive knowledge about various types of elementary school programs. At this time our best opportunity to gain this essential information lies in designing and conducting a coherent set of follow-up studies of the Follow Through Program. Through such comprehensive follow-up studies we should find answers to fill critical gaps in our current understanding of elementary school programs.

References

Becker, W. C., & Gersten, R. (1982). A follow-up of Follow Through: Meta-analysis of the later effects of the Direct Instruction Model. *American Educational Research Journal*, 19, 1, 75-92.

Bereiter, C. & Kurland, M. (1981-82). A constructive look at Follow Through results. *Interchange*, 12, 1-22.

Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W.S., Epstein, A. S., & Weikart, D. P. (1984). Changed lives: The effects of the Perry Preschool program through age 19. *Monographs of the High/Scope Educational Research Foundation*, 8.

Clark, A. (1975). Follow-up on Follow Through. *Measurement and Evaluation in Guidance*. 7, (4), 234-238.

Cloud, K., Rentfrow, R. K., Hildebrandt, J., Abrams, B., & DeCausey, B. (1980). *Later school success of Follow Through graduates in Lakewood, New Jersey public schools*. Tucson: University of Arizona Center for Educational Research and Development.

Egbert, R. (1973). *Planned variation in Follow Through*. Paper delivered before the Brookings Institution Panel on Social Experimentation, Washington, DC.

Egbert, R., & England, M. (1991). *Follow Through: A brief history*. (Paper prepared for a meeting considering long term follow-up studies of Follow Through. Mimeographed.)

Gersten, R., Carnine, D., & Keating, T. (1984). *Lasting impact of the Direct Instruction Follow Through Program: Preliminary findings of a longitudinal study of 1500 students*. Paper presented as part of a symposium at the American Research Association Annual Meeting, New Orleans.

Haney, W. (1977). The Follow Through Experiment: Summary of an analysis of major evaluation reports. *Curriculum Inquiry*, 7, (3), 227-257.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. (1978). No simple answer: Critique of the "Follow Through" evaluation. *Harvard Educational Review*. 48, 128-160.

Kennedy, M. M. (1978). Findings from the Follow Through planned variation study. *Educational Researcher*, 7, 3-11.

Lazar, I., Darlington, R., Murray, H., Royce, J., & Snipper, A. (1982). Lasting effects of early education. *Monographs of The Society for Research in Child Development*, 47(1-2, Serial No. 194).

National Commission on Excellence in Education. (1983). *A nation at risk*. Government Printing Office. Washington, DC.

Oden, S. (1991, February). *Feasibility of a followup study of Follow Through graduates: Strategies for subject identification, location, and interview former staff, students, and parents*. (Paper prepared for a meeting considering long term follow-up studies of Follow Through. Mimeographed.)

Olmstead, P. (1987). Mimeographed. Ypsilanti, MI: High/Scope Educational Research Foundation.

President's Education Summit. (1990). National Educational Goals.

Rhine, W. R. (Ed.). (1981). *Making schools more effective: New directions from Follow Through*. New York: Academic Press.

Sells, S. B., Loftus, J. J., & Herbert. (1941). Evaluative studies of the activity program in New York City public schools: A preliminary report. *Journal of Experimental Education*. 9, 310-322.

BEST COPY AVAILABLE

Stallings, J. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. *Monographs of the Society for Research in Child Development*, 40(7-8), 1-133.

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1977). *Education as experimentation: A planned variation model, Vols. IVA-D. An Evaluation of Follow Through.* Cambridge, MA: Abt Associates.

Walberg, H. (1986). Syntheses of research on teaching. In M.C. Wittrock (Ed.) *Handbook of research on teaching.* New York: Macmillan.

Wang, M., & Walberg, H. (Undated). Mimeographed.

Wang, M. C., & Ramp, E. (Eds.). (1987). *The National Follow Through Program: Design, implementation, and effects.* (Final Project Report, Vols. 1 and 2). Philadelphia, PA: Temple University Center for Research in Human Development and Education.

Werner, E. E., & Smith, R. S. (1982). *Vulnerable but invincible.* New York: McGraw-Hill.

Wolf, M., & Stein, A. (1966). *Study I. Six months later, a comparison of children who had Head Start summer 1965, with their classmates in kindergarten* (a case study of kindergartens in four public elementary schools, New York City). Paper presented at the meeting of the American Psychological Association.